



JP2000339406

Biblio

Page 1

esp@cenet

**SLIP RECOGNIZING METHOD**

Patent Number: JP2000339406

Publication date: 2000-12-08

Inventor(s): OBARA KATSUTOSHI; EGUCHI SHINICHI; CHIBA NOBUKAZU;
KANEMOTO KOICHI; YABUKI MASANORI; KATSUMATA YUTAKA

Applicant(s):: FUJITSU LTD

Requested

Patent: ☐ JP2000339406 (JP00339406)

Application

Number: JP19990149154 19990528

Priority Number

(s):

IPC

Classification: G06K9/20 ; G06F19/00 ; G06T7/00 ; G06T7/60

EC

Classification:

Equivalents:

Abstract

PROBLEM TO BE SOLVED: To obtain a method for automatically recognizing defining information of each item of a slip by analyzing the table structure of the slip from the physical arrangement of the smallest rectangle consisting of vertical and horizontal ruled lines so as to determine an attribute of the smallest rectangle.

SOLUTION: A reader 1000 optically reads a slip and a processor 1001 analyzes the slip recognize characters. An entry wen-yen dictionary 1003 in a memory 1002 stores a previously fixed entry wen-yen and its image data and a general dictionary 1004 stores a dictionary by each character recognizing category/character font in order to recognize the characters in a data part. In a CPU 1001, a ruled line extracting part 1005 extracts the ruled lines of the slip and the smallest rectangle consisting of the ruled lines and an analyzing part 1006 analyzes the arrangement of the extracted smallest rectangle to analyze table structure. An attribute deciding part 1007 decides the attribute of each smallest rectangle and a character recognizing part 1008 recognizes the characters in accordance with its attribute.

Data supplied from the esp@cenet database - I2

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-339406
(P2000-339406A)

(43) 公開日 平成12年12月8日 (2000.12.8)

(51) Int.Cl. ⁷	識別記号	F I	テマコード* (参考)
G 0 6 K 9/20	3 4 0	G 0 6 K 9/20	3 4 0 C 5 B 0 2 9
G 0 6 F 19/00		G 0 6 F 15/22	3 1 0 A 5 L 0 9 6
G 0 6 T 7/00		15/70	3 3 0 Q
	7/60		3 5 0 L

審査請求 未請求 請求項の数12 O L (全 57 頁)

(21) 出願番号 特願平11-149154

(22) 出願日 平成11年5月28日 (1999.5.28)

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 小原 勝利

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(72) 発明者 江口 真一

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(74) 代理人 100094514

弁理士 林 恒徳 (外1名)

最終頁に続く

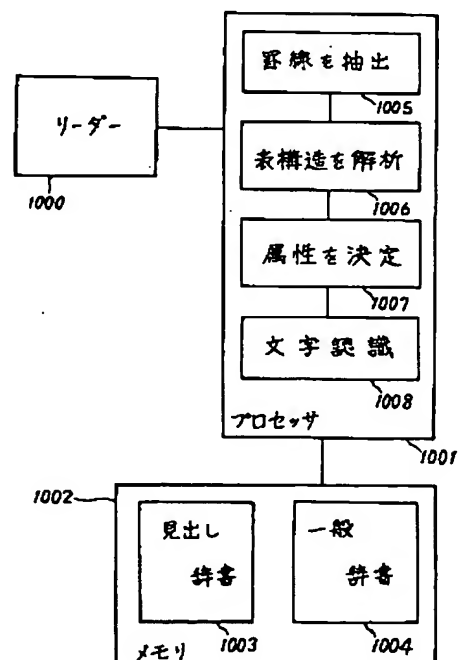
(54) 【発明の名称】 帳票認識方法

(57) 【要約】

【課題】 帳票のフォーマットを自動認識する帳票認識方法に関し、各最小矩形の並びからフォーマットを自動決定する。

【解決手段】 帳票上の最小矩形を抽出し、最小矩形の位置関係を判定する。この位置関係から最小矩形の属性を決定する。最小矩形の属性により、見出し部分とデータ部分とに区別して、データ部分の文字列を認識する。

ブロック図



【特許請求の範囲】

【請求項 1】 帳票の表構造を認識する帳票認識方法において、

前記帳票のイメージから前記帳票の縦罫線、横罫線及び罫線で構成される最小矩形を抽出するステップと、
前記帳票上の前記最小矩形の物理的並びから前記帳票の表構造を解析するステップと、
前記解析した表構造から前記最小矩形の属性を決定するステップとを有することを特徴とする帳票認識方法。

【請求項 2】 請求項 1 の帳票認識方法において、
前記表構造を解析するステップは、
前記最小矩形の位置関係から前記最小矩形の行方向の関係を抽出するステップと、
前記最小矩形の位置関係から前記最小矩形の列方向の関係を抽出するステップとを有することを特徴とする帳票認識方法。

【請求項 3】 請求項 2 の帳票認識方法において、
前記行方向の関係を抽出するステップは、
前記最小矩形の接続関係から前記最小矩形を各行を構成する最小矩形に分類するステップと、
前記連続して、且つ同じ行構造を持つ行をブロックにまとめるステップとを有し、
前記列方向の関係を抽出するステップは、
上下に接続するブロック間で同じ列構造を持つブロックの関係を抽出するステップを有するステップを有することを特徴とする帳票認識方法。

【請求項 4】 請求項 3 の帳票認識方法において、
前記属性を決定するステップは、
属する行数が最大であるブロックをデータ部分として決定し、そのブロックの上及び下に位置するブロックを見出し部分として決定するステップを含むことを特徴とする帳票認識方法。

【請求項 5】 請求項 4 の帳票認識方法において、
前記属性を決定するステップは、
前記見出し部分の文字を文字認識して、前記見出し部分の属性を決定するステップと、
前記見出し部分の属性から前記データ部分の属性を決定するステップとを有することを特徴とする帳票認識方法。

【請求項 6】 請求項 1 の帳票認識方法において、
前記表構造を解析するステップは、
前記最小矩形の接続関係から、接続する最小矩形をグループに分類するステップと、
前記グループを表の構成要素単位に分類するステップと、
前記表の構成要素単位に、前記最小矩形間の行と列の関係を解析するステップとを有することを特徴とする帳票認識方法。

【請求項 7】 請求項 6 の帳票認識方法において、
前記最小矩形間の行と列の関係を解析するステップは、

前記最小矩形を、行方向の関係に解析するステップと、
前記最小矩形を、列方向の関係に解析するステップとを有することを特徴とする帳票認識方法。

【請求項 8】 請求項 7 の帳票認識方法において、
前記行方向の関係に解析するステップは、
最小矩形が構成する行の入れ子構造を解析して、前記最小矩形に、入れ子情報を設定するステップを有することを特徴とする帳票認識方法。

【請求項 9】 請求項 6 の帳票認識方法において、
前記解析するステップは、
前記各構成要素の表構造を抽出するステップと、
同一の表構造を有する構成要素を統合するステップとを有することを特徴とする帳票認識方法。

【請求項 10】 請求項 6 の帳票認識方法において、
前記解析するステップは、
前記各構成要素の表構造を抽出するステップと、
前記構成要素の表構造から、前記構成要素内の罫線を回復するステップを有することを特徴とする帳票認識方法。

【請求項 11】 請求項 7 の帳票認識方法において、
前記行方向の関係を解析するステップは、
前記各行の構造から、前記行に付与された消し線を検出するステップと、
前記消し線を除去して、各行の関係を決定するステップとを有することを特徴とする帳票認識方法。

【請求項 12】 請求項 6 の帳票認識方法において、
前記解析するステップは、
同じ大きさの最小矩形が連続することを検出するステップと、
連続した最小矩形を 1 つの最小矩形に統合するステップとを有することを特徴とする帳票認識方法。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】 本発明は、帳票上の文字を認識するため、帳票のフォーマットを認識するための帳票認識方法に関し、特に、表を持つ帳票の表構造を自動解析して、表を構成する各項目の属性を決定する帳票認識方法に関する。

【0002】 データの自動入力のため、帳票上の文字を認識する文字認識方法が用いられている。この文字認識においては、帳票上の指定された位置のイメージを取り出し、文字認識する。

【0003】 帳票上の文字を識別するには、その文字のデータ名（フィールド認識子）、文字種類等の文字の属性を定義する必要がある。

【0004】

【従来の技術】 図 87 は、従来技術の説明図である。

【0005】 図 87 は、帳票として振込依頼書を示す。この振込依頼書は、「振込先」、「銀行名」、「支店名」等の項目がある。このような項目を自動認識するに

は、各項目のフィールドID（データ名称）、開始位置、終了位置、桁数、認識カテゴリ（文字種類）等の属性を定義する必要がある。従来、この定義情報は、各帳票毎に、登録していた。この登録には、オペレータが、各帳票毎に、入力していた。

【0006】

【発明が解決しようとする課題】しかしながら、従来技術では、定義情報を事前に登録する必要があった。このため、予め定義情報を登録した帳票しか、文字認識できない。例えば、金融機関において、振込の自動入力を行う場合に、企業が作成する振込依頼書のフォーマットは、様々である。このような各帳票の定義情報を予め作成しておくことは、手間がかかるという問題がある。

【0007】又、帳票の定義情報を登録したとしても、その帳票のフォーマットが変更になった場合に、定義情報を変更する必要があるという問題もある。

【0008】本発明の目的は、帳票の各項目の定義情報を自動認識する帳票認識方法を提供するにある。

【0009】本発明の他の目的は、帳票の各項目の並びを解析して、各項目の属性を決定する帳票認識方法を提供することにある。

【0010】本発明の更に他の目的は、帳票の最小矩形の属性を決定して、帳票のデータ部分を文字認識するための帳票認識方法を提供することにある。

【0011】

【課題を解決するための手段】本発明は、帳票の表構造を認識する帳票認識方法である。その方法は、帳票のイメージから前記帳票の縦罫線、横罫線及び罫線で構成される最小矩形を抽出するステップと、前記帳票上の前記最小矩形の物理的並びから前記帳票の表構造を解析するステップと、前記解析した表構造から前記最小矩形の属性を決定するステップとを有する。

【0012】本発明では、表構造の最小矩形を抽出し、最小矩形の物理的並びを検出する。この最小矩形の物理的並びから帳票の表構造を解析して、最小矩形の属性を決定する。最小矩形の物理的並びを検出して、最小矩形の属性を決定するので、帳票の表構造を自動解析することができる。

【0013】このため、帳票の各項目を、事前に登録する手間を省くことができる。又、帳票のフォーマットが変更されても、定義情報を人手により変更する手間を省くことができる。

【0014】本発明の他の形態は、前記表構造を解析するステップは、前記最小矩形の位置関係から前記最小矩形の行方向の関係を抽出するステップと、前記最小矩形の位置関係から前記最小矩形の列方向の関係を抽出するステップとを有する。

【0015】更に、本発明の他の形態は、前記行方向の関係を抽出するステップは、前記最小矩形の接続関係から前記最小矩形を各行を構成する最小矩形に分類するス

テップと、前記連続して、且つ同じ行構造を持つ行をブロックにまとめるステップとを有し、前記列方向の関係を抽出するステップは、上下に接続するブロック間で同じ列構造を持つブロックの関係を抽出するステップを有するステップを有する。

【0016】本発明の更に他の形態は、前記属性を決定するステップは、属する行数が最大であるブロックをデータ部分として決定し、そのブロックの上及び下に位置するブロックを見出し部分として決定するステップを含む。

【0017】本発明の更に他の形態は、前記属性を決定するステップは、前記見出し部分の文字を文字認識して、前記見出し部分の属性を決定するステップと、前記見出し部分の属性から前記データ部分の属性を決定するステップとを有する。

【0018】本発明の更に他の形態は、前記表構造を解析するステップは、前記最小矩形の接続関係から、接続する最小矩形をグループに分類するステップと、前記グループを表の構成要素単位に分類するステップと、前記表の構成要素単位に、前記最小矩形間の行と列の関係を解析するステップとを有する。

【0019】本発明の更に他の形態は、前記最小矩形間の行と列の関係を解析するステップは、前記最小矩形を、行方向の关系到解析するステップと、前記最小矩形を、列方向の关系到解析するステップとを有する。

【0020】本発明の更に他の形態は、前記行方向の关系到解析するステップは、最小矩形が構成する行の入れ子構造を解析して、前記最小矩形に、入れ子情報を設定するステップを有する。

【0021】本発明の更に他の形態は、前記解析するステップは、前記各構成要素の表構造を抽出するステップと、同一の表構造を有する構成要素を統合するステップとを有する。

【0022】本発明の更に他の形態は、前記解析するステップは、前記各構成要素の表構造を抽出するステップと、前記構成要素の表構造から、前記構成要素内の罫線を回復するステップを有する。

【0023】本発明の更に他の形態は、前記行方向の关系到解析するステップは、前記各行の構造から、前記行に付与された消し線を検出するステップと、前記消し線を除去して、各行の関係を決定するステップとを有する。

【0024】本発明の更に他の形態は、前記解析するステップは、同じ大きさの最小矩形が連続することを検出するステップと、連続した最小矩形を1つの最小矩形に統合するステップとを有する。

【0025】

【発明の実施の形態】図1は、本発明の一実施の形態の帳票認識装置のブロック図、図2は、本発明の一実施の形態に用いられる第1の帳票例の説明図、図3は、本発

明の一実施の形態に用いられる第2の帳票例の説明図である。

【0026】図1に示すように、リーダー1000は、帳票を光学的に読み取る。プロセッサ(CPU)1001は、帳票を解析して、文字認識を行う。メモリ1002は、見出し文言辞書1003、一般辞書1004を格納する。見出し文言辞書1003は、予め定められた見出し文言とそのイメージデータを格納する。一般辞書1004は、データ部分の文字認識のため、文字認識カテゴリー／文字フォント別の辞書を格納する。

【0027】CPU1001は、帳票の罫線と、罫線により構成される最小矩形を抽出する罫線抽出部1005と、抽出した最小矩形の並びを解析して、表構造を解析する解析部1006と、解析した表構造から、各最小矩形の属性を決定する属性決定部1007と、属性に従い、文字認識を行う文字認識部1008とを有する。尚、後述するように、これら各部は、CPU1001が行う処理をブロックで表現したものである。

【0028】図2は、解析される第1の帳票例を示す。この帳票は、振込依頼書である。図2の第1の帳票は、依頼人、依頼日、振込予定日、振込先(銀行名、支店名)、科目、口座番号、受取人名、金額、小計、合計を記入するものである。この帳票において、〔数字〕は、横方向に延びる横罫線を、〔数字〕は、縦方向に延びる縦罫線を、数字は、縦罫線及び横罫線で囲まれる最小矩形(以降、罫線項目と呼ぶ)を示す。従って、図2の帳票は、20の横罫線(1)～(20)と、13の縦罫線〔1〕～〔13〕と、176の罫線項目1～176とで構成されている。

【0029】図3は、解析される第2の帳票例を示す。この帳票も、振込依頼書である。図3の第2の帳票は、依頼人、銀行名、支店名、科目、口座番号、受取人名、金額、小計、合計、適用、備考を記入するものである。この帳票においても、〔数字〕は、横方向に延びる横罫線を、〔数字〕は、縦方向に延びる縦罫線を、数字は、縦罫線及び横罫線で囲まれる最小矩形(以降、罫線項目と呼ぶ)を示す。従って、図3の帳票は、30の横罫線(1)～(30)と、20の縦罫線〔1〕～〔20〕と、171の罫線項目1～171とで構成されている。この帳票は、表の部分が2つに分割されている。

【0030】本発明は、このような表構造を有する帳票の各罫線項目の種類等を自動解析するものである。

【0031】図4及び図5は、本発明の一実施の形態の全体フロー図、図6乃至図25は、その処理の説明図である。

【0032】(S1) CPU1001は、帳票のイメージデータから帳票上の横罫線と縦罫線と、横罫線と縦罫線に囲まれた最小矩形(罫線項目)を抽出する。図6(A)の示すように、イメージデータの黒ドットが連続する数を数え、一定数以上のドットが連続した場合に、

横罫線又は縦罫線と判断する。図6(B)は、横罫線と縦罫線との例を示す。罫線は、物理的な座標として抽出され、縦及び横罫線を開始座標と終了座標で示す。罫線項目は、図6(C)に示すように、2本ずつの縦罫線及び横罫線で囲まれる最小矩形を抽出したものである。この罫線項目も、開始座標と終了座標で示す。

【0033】次に、CPU1001は、イメージデータの帳票の斜度を算出し、斜度に合わせて罫線項目、横罫線、縦罫線の各開始座標及び終了座標を補正する。

【0034】(S2) CPU1001は、図7に示すように、各罫線項目を、接続する位置関係にあるもの同志が同じグループになるように、分類する。各罫線項目の座標を参照して、上下又は左右に接続する位置関係にあるもの同志を同じグループに分類する。図7に示すように、図2の第1の帳票は、3つのグループに分類される。図8は、接続関係の態様を示したものである。図8に示すように、罫線項目が接続しているかを判断するには、横方向又は縦方向に完全の接続している場合のみならず、罫線項目間の距離が、横方向又は縦方向で、所定の閾値の範囲内にあるものも、接続していると判断する。

【0035】(S3) 次に、CPU1001は、グループを表の構成要素単位に分類する。罫線項目を分類したグループを、更にグループ毎に、表構造の構成単位に細分化する。図9に示すように、構成単位とは、グループの主となる表の部分、上側に突出した部分(上凸部)、下側に突出した部分(下凸部)からなる。このため、各罫線が、座標からどのグループ内に位置するか(どのグループを構成しているか)を判定する。そして、それぞれのグループ単位に、2本ずつの縦罫線及び横罫線で矩形(正方形、長方形)を構成できるものを探す。そして、罫線で構成される矩形の中で、最も大きいものをグループ内での最大矩形とする。グループ単位に、最大矩形に含まれない位置に存在する罫線項目を探索し、最大矩形の上側に位置するものを、上凸部、下側に位置するものを下凸部と分類する。そして、最大矩形に含まれる罫線項目を主表部に分類する。以降、この細分化した単位をグループと呼ぶ。

【0036】図9の帳票で、最大矩形を求めた場合、最大矩形は、横罫線a、横罫線b、縦罫線c、縦罫線dで囲まれる範囲となる。従って、主表部は、罫線項目3～26、上凸部は、罫線項目1～2、下凸部は、罫線項目27～32となる。

【0037】(S4) CPU1001は、表の構成要素単位のグループを論理行に分類する。グループ単位に、構成する罫線項目を、行方向(帳票の横方向)に論理的接続を持つもの同志を纏めた論理行に、分類する。ここで言う論理的な接続とは、同幅接続という概念により接続する状態にあるものをいう。

【0038】同幅接続の関係にある罫線項目は、同一の

論理行を構成すると判断する。同幅接続について、図10(A)乃至図10(D)により、説明する。

【0039】図10(A)に示すように、罫線項目が、同じ行幅で隣接し、且つ互いに罫線項目の構成線分を共有する場合、参照中の罫線項目が同一の論理行を構成すると判断する。図10(A)では、罫線項目Aと罫線項目Bでは、線分a4と線分b1とが同一の線分であり、線分の長さも同一であることから、2つの罫線項目は、接続状態にあると判断する。

【0040】図10(B)に示すように、罫線項目同志が近接し、同じ行幅を持つ場合、参照中の罫線項目が同一の論理行を構成すると判断する。図10(B)に示すように、罫線項目Aと罫線項目Bとの距離dが、所定の閾値以下で、線分a4と線分b2が同じ長さである場合、2つの罫線項目は、接続状態にあると判断する。

【0041】図10(C)に示すように、1つの罫線項目に対し、複数の罫線項目が同じ幅で隣接し、互いに罫線項目の構成線分を共有する場合、参照中の罫線項目が同一の論理行を構成すると判断する。図10(C)では、罫線項目Bの線分b2と罫線項目Cの線分c2との和が、罫線項目Aの線分a4の長さと同じであることから、3つの罫線項目は、接続状態にあると判断する。

【0042】図10(D)に示すように、各罫線項目が近接し、1つの罫線項目の行幅と、複数の罫線項目の行幅の和が同じ場合には、参照中の罫線項目が同一の論理行を構成すると判断する。図10(D)では、罫線項目Bと罫線項目Cとの距離d2が、閾値以下であり、且つ罫線項目Aと罫線項目B、Cとの距離d1が閾値以下である。罫線項目Bの線分b2と罫線項目Cの線分c2との和が、罫線項目Aの線分a4の長さにはほぼ等しいことから、3つの罫線項目は、接続状態にあると判断する。

【0043】このようにして、各グループを論理行に分類する。図11に示すように、図2の帳票のグループ3は、11の論理行に分類される。

【0044】次に、行方向の入れ子構造部を解析する。複数の罫線項目から構成される矩形が論理的な意味をもつものを、入れ子構造という。図12(A)に示すように、「振込先」の項目に対して、「銀行名」及び「支店名」の項目は、関係するから、この部分を入れ子構造と定義する。図12(B)に示すように、図12(A)の場合には、罫線項目1(振込先)が、罫線項目2(銀行名)の親である。従って、罫線項目1と2に、親子関係を設定する。罫線項目2(銀行名)と罫線項目3(支店名)とは、兄弟関係(同レベル関係)である。従って、罫線項目2と3に、兄弟関係を設定する。

【0045】又、図3の第2の帳票例のように、レイアウトの都合上、表構成部を左右に分割している場合がある。これを、表の構造の論理的意味から判断すると、表構造を決定する上で、1つの表構造としてまとめる必要がある。このため、図13に示すように、左右に並んで

位置するグループが複数存在する場合に、それぞれのグループを構成する論理行構造を判定する。それぞれの論理行によるグループの構造が同じ場合に、表構造を統合する。

【0046】更に、左右に分割されている表の一部が掠れ等により消失した場合に、その部分に罫線を作成する。前述の左右に分離された表を統合する場合に、一方の表のどれかの罫線が掠れ等で消失している場合がある。図14に示すように、罫線が消失している部分以外は、2つのグループの論理行構造が一致していることから、他方の論理行構造と一方の論理行構造とを合わせるように、罫線を新規に作成する。これにより、生成した罫線を含む罫線項目を生成する。このように、罫線を作成した後、前述の表構造の統合処理を行う。

【0047】(S5)次に、連続する同じ構造の論理行をブロック単位に纏める。1グループ内で上下に隣り合う論理行同志の構造を比較する。そして、同じ構造を持つ論理行同志を1ブロックに分類する。論理行の比較要素として、構成する罫線項目の数、論理行中の罫線項目の配置、論理行の高さと幅がある。これらを比較して、同じ構造を持つか判定する。

【0048】図15の帳票例では、1つのグループの7つの論理行が、3つのブロック1~3に分類される。ブロック1は、論理行1で構成され、ブロック2は、論理行2~6で構成され、ブロック3は、論理行7で構成される。

【0049】又、異なる構造のブロック間を同一のブロック構造に変換する。即ち、表の一部の罫線が掠れ等で消失している場合に、他の論理行構造から消失した罫線を新規に生成する。それに伴い罫線項目を生成する。

【0050】図16の表の例で説明すると、罫線が消失し、ブロック1~5に分類されている。しかし、ブロック2とブロック4は、同一の構造を持つブロックである。従って、ブロック3に、罫線を作成する。これにより、ブロック2~ブロック4は、同一の構造となる。これにより、ブロックを3つに統合する。

【0051】次に、消し線により分割された論理行を統合する。金融機関で扱う給与振込帳票等では、処理の対象としないことを明確にするため、消し線で取引部分を消去する場合がある。この場合、消し線は罫線として抽出されるため、解析結果は、不正確になる。従って、消し線を除いて、表構造の解析を行う必要がある。ブロックに分類された状態では、消し線で引かれた部分のブロックは、次の特徴がある。

【0052】第1に、上下に別のブロックが存在する。第2に、上側ブロックと下側ブロックは、同じ構造を持つ。この特徴を持つブロックが存在する場合に、当ブロック内で、消し線による罫線を除いたブロックを作成する。そして、このブロックが、上側ブロックと同一構造となる場合に、消し線が引かれたことにより、ブロック

構造が変化したものと判断する。そして、消し線を消去し、これに伴って、罫線項目、論理行、及びブロックを再作成する。

【0053】図17に示すように、ブロック3とブロック5に消し線が存在すると、ブロックは7つに分類される。ここで、ブロック2とブロック4は、同じ構造であり、ブロック3の消し線を消去すると、ブロック2と同じ構造となる。又、ブロック4～6についても、同様である。従って、消し線を消去し、罫線項目、論理行を再作成する。それに伴い、ブロックも作成し直す。これにより、3ブロックに再分類される。

【0054】(S6)次に、論理列構造を判定する。表構造を持つものは、文言固定部と、文言可変部とからなる。文言固定部は、表の見出し部分などの帳票の固定部分である。文言可変部は、表のデータ部分などの帳票の可変部分である。論理列構造の判定は、文言固定部と文言可変部との論理列構造を判定する。即ち、上下に接続するブロックの間に、同じ列構造を持つかを判定する。そして、同じ列構造を持つ場合、それらブロックを論理列の単位で統合する。例えば、図18(A)の例では、論理列は、6つに分類される。

【0055】この場合、図18(B)に示すように、文言可変部のブロック2では、「あいう銀行」と「かきく支店」の2列の構造であるが、文言固定部のブロック1では、1列の構造となっている。この場合、ブロック1の1列(振込先)がブロック2の2列の見出しとなっているため、表構造としてみた場合には、ブロック2の2列を1列とみなす。

【0056】同様に、図18(C)に示すように、文言固定部のブロック1では、「銀行名」と「支店名」の2列の構造であるが、文言可変部のブロック2では、1列の構造となっている。この場合、ブロック1の2列(銀行名、支店名)がブロック2の1列の見出しとなっているため、表構造としてみた場合には、ブロック1の2列を1列とみなす。

【0057】次に、列方向の論理的な入れ子構造を抽出する。ブロック間での論理列構造の判定を行った結果、図18(B)及び図18(C)のように、一方のブロックが、複数列からなり、他方のブロックが1列の論理列構造を持つ場合、これらを入れ子構造として表現する。入れ子構造を構成する罫線項目は、纏めて1つの論理的な意味を持たせる。

【0058】図19(A)は、図18(B)の場合の入れ子構造を示す。図19(A)に示すように、罫線項目2と罫線項目3が兄弟関係(同じレベル)にある。両罫線項目をあわせて、罫線項目1に対しての列を構成する罫線項目とする。図19(B)は、図18(C)の場合の入れ子構造を示す。図19(B)に示すように、罫線項目1と罫線項目2が兄弟関係(同じレベル)にある。両罫線項目をあわせて、罫線項目3に対しての列を構成

する罫線項目とする。

【0059】更に、1桁罫線項目部分、3桁罫線項目部分の入れ子構造を解析する。図20(A)及び図20(B)に示すように、帳票の表構造の中で、金額欄、手数料欄がある場合に、1桁の罫線項目又は3桁の罫線項目が連続する。1桁罫線項目は、1罫線項目内に1桁の数字が記入又は印刷されるものをいう。又、3桁罫線項目は、1罫線項目に3桁の数字が記入又は印刷されるものをいう。

【0060】これらは、合わせて、1つの意味を持つ。これらを論理的に1つの罫線項目に纏める。このため、予め1桁罫線項目又は3桁罫線項目の縦サイズと横サイズを定義しておく。これらサイズと一致する罫線項目が任意数並んだ場合には、これらを入れ子構造として、論理的に1つの罫線項目としてまとめる。

【0061】このような結果を利用して、文字認識処理では、連続する罫線項目を1つの文字列として扱うことができる。例えば、図20(A)のa部分の罫線項目内を文字認識した結果、b部分の結果を得た場合に、罫線項目が入れ子構造となっていることから、文字認識結果をc部分で示したように結合して、1つの文字列として取り扱う。図20(B)に示す、3桁罫線項目の場合も同様である。

【0062】次に、特殊な構造を持つ矩形の入れ子構造を解析する。帳票上の主表以外の箇所には、特別な構造を持つものが多い。これらについては、構造毎に入れ子構造解析処理を行う。図21(A)では、罫線項目1に対し、罫線項目2が親子の関係にある。罫線項目2に対し、罫線項目4が親子の関係にある。罫線項目3に対し、罫線項目5が親子の関係にある。罫線項目2に対し、罫線項目3が兄弟の関係にある。罫線項目4に対し、罫線項目5が兄弟の関係にある。図21(B)では、罫線項目1に対し、罫線項目2が親子の関係にある。罫線項目2に対し、罫線項目4が親子の関係にある。罫線項目3に対し、罫線項目5が親子の関係にある。罫線項目2に対し、罫線項目3が兄弟の関係にある。罫線項目4に対し、罫線項目5が兄弟の関係にある。前述した行方向及び列方向の入れ子構造の解析処理と同様にして、入れ子構造を解析する。

【0063】(S7)次に、表構造部のブロックに属性を付与する。帳票には、見出し部と、見出し部分に対応するデータ部とが存在する。又、帳票によっては、データを纏める部分(例えば、データ部の金額合計欄等)が存在する。ここでは、表構造部を構成するブロックに、属性を付与して、各ブロックに意味を持たせる。即ち、図22に示すように、表のデータを記入又は印刷する部分をボディ部と定義し、表を構成するブロック中で最も論理行数が多いブロックに、ボディ部属性を付与する。表の見出しを記入又は印刷する部分をヘッダ部と定義し、ボディ部の上側に位置するブロックに、ヘッダ部属

性を付与する。表のデータ部分を取りまとめるデータ（例えば、合計欄等）を記入又は印刷する部分をフッタ部と定義し、ボディ部の下側に位置するブロックにフッタ部属性を付与する。

【0064】（S8）次に、見出しを構成する罫線項目に、見出しフラグを設定する。表構造の形式から以下に示すルールに従って、見出しにあたる罫線項目に見出しフラグを設定する。見出しフラグを設定する罫線項目は、図23に示すように、次の通りである。

- 【0065】・ヘッダブロックを構成する罫線項目
- ・ボディブロックの最上位論理行を構成する罫線項目
- ・フッタブロックを構成する罫線項目の内、最も右に位置するものを除いた罫線項目
- ・下凸部の論理行を構成する罫線項目の内、最も右に位置するものを除いた罫線項目
- ・表以外の矩形の論理行を構成する罫線項目の場合、入れ子構造の罫線項目の内、最も右下に位置するものを除いた罫線項目

この見出し部を抽出しておく、と、使用される文字を制限でき、その文字（文言）に特化した文字認識処理を行うことができる。従って、高精度に見出し部分の文言を認識することができる。

【0066】（S9）次に、見出しの種類を決定する。見出しフラグが設定された罫線項目内に存在する文字列の文字認識を行う。図25に示すように、この見出し部の文字認識で使用する認識辞書には、見出し文言辞書1003（図1参照）が使用される。見出し文言辞書1003は、見出しとして存在する文言（振込先等）と、そのイメージとが登録されたものである。従って、一般の文字認識辞書を使用した場合よりも高い精度での文字認識が可能となる。

【0067】この認識結果から得られた見出し名称をキーとして、図24に示す見出しテーブルを参照する。図24に示すように、見出しテーブルには、見出し名称に対応する属性情報（文字認識カテゴリ、文字フォント等）が登録されている。そして、その見出し名称に一致するレコード（属性情報）を抽出する。

【0068】（S10）次に、見出しに対応する罫線領域の認識属性を決定する。このため、見出しが見つかった罫線項目を含む論理列情報を参照し、その見出しに対応するデータ部分の罫線項目を見つける。即ち、論理列を構成する罫線項目のうち、見出し以外の罫線項目に対して、文字認識カテゴリ、文字フォント等の認識特性を、見出しの罫線項目から複写する。

【0069】（S11）次に、見出し以外の領域の文字認識を行う。見出しの罫線項目以外の罫線項目内に存在する文字列を、一般の文字認識辞書1004（図1参照）を使用して、文字認識する。この時、文字認識辞書1004は、漢字、数字等の各認識カテゴリに対し、別々に設けられている。従って、前述したデータ部分の認

識属性に従い、一般辞書から認識辞書を選択して、文字認識する。このため、データ部分も高い精度で文字認識できる。

【0070】更に、得られた認識結果を認識結果通知領域に設定する。これにより、データ部分のデータ名称、データ数、文字認識結果が得られる。

【0071】このようにして、各罫線項目の行と列の関係を判定し、見出し部とそれに対応するデータ部とに分類する。そして、見出し部分を見出し認識辞書を用いて文字認識して、見出し名称を得る。論理列の構造の関係から見出し部分に対応したデータ部分のデータ名称が判明する。このため、帳票の各項目のフォーマットを予め定義しておく必要がない。

【0072】更に、見出し名称から認識属性を見出しテーブルにより得る。これに従い、データ部分を文字認識する。このため、帳票の表構造を解析することができ、データ部分の文字認識結果を精度を高めることができる。

【0073】次に、前述のステップS1～S11の各処理について、詳細に説明する。ここで、図26は、各処理に使用されるテーブルの関係図である。図26に示すように、横罫線テーブル1105は、抽出した横罫線の開始座標と終了座標を格納する。縦罫線テーブル1106は、抽出した縦罫線の開始座標と終了座標を格納する。罫線項目情報テーブル1104は、抽出した罫線項目（最小矩形）の開始座標、終了座標と、属するグループ番号、ブロック番号等を格納する。グループ情報テーブル1100は、分類された各グループの先頭罫線項目アドレス、先頭ブロック番号等を格納する。

【0074】ブロック情報テーブル1101は、抽出された各ブロックの先頭罫線項目アドレス、先頭論理行アドレス等を格納する。論理行情報テーブル1102は、抽出された各論理行の先頭罫線項目アドレス、ブロック番号等を格納する。論理列情報テーブル1103は、抽出された各論理列の先頭罫線項目アドレス等を格納する。

【0075】（a）罫線抽出処理の説明

図27は、横罫線情報テーブルの説明図、図28は、縦罫線情報テーブルの説明図、図29は、罫線項目テーブルの説明図である。

【0076】横罫線の抽出は、周知のように、一定数以上の黒画素が横方向に連続する黒画素群を抽出する。そして、図27に示すように、各抽出した横罫線に番号を付し、その開始座標及び終了座標を、横罫線情報テーブル11005に格納する。

【0077】縦罫線の抽出も、周知のように、一定数以上の黒画素が縦方向に連続する黒画素群を抽出する。そして、図28に示すように、各抽出した縦罫線に番号を付し、その開始座標及び終了座標を、縦罫線情報テーブル11006に格納する。

【0078】最小矩形（野線項目）の抽出も、周知のように、2つの横野線と2つの縦野線に囲まれた最小矩形を抽出する。そして、各抽出した野線項目に番号を付す。図29に示すように、野線項目情報テーブル1104は、各野線項目の開始座標SC、終了座標EC、上側横野線番号ULN、下側横野線番号LLN、左側縦野線番号LVN、右側縦野線番号RVNを格納する。尚、野線項目情報テーブル1104は、グループ番号GN、ブロック番号BN、行番号LN、列番号RO、上側野線項目URN、下側野線項目LRN、左側野線項目ERN、右側野線項目RRN、入れ子情報INの各欄を有する。

$$\tan \theta = (Y_e - Y_s) / (X_e - X_s)$$

そして、ステップS20に戻る。

【0083】(S22) ステップS20で、全横野線を参照した場合には、斜度の平均値を計算する。斜度の平均値は、(斜度値の合計) / (横野線の本数) により計算される。そして、終了する。

【0084】次に、座標補正処理について、説明する。図31は、座標補正処理フロー図である。

【0085】(S25) 野線項目テーブル1104の各

$$X = X_s + Y_s \times \tan \theta$$

$$Y = Y_s - X_s \times \tan \theta$$

そして、野線項目の右下座標（終了座標） X_e 、 Y_e を、斜度 $\tan \theta$ に応じて補正する。補正後の座標X、

$$X = X_e + Y_e \times \tan \theta$$

$$Y = Y_e - X_e \times \tan \theta$$

そして、ステップS25に戻る。

【0089】(S27) 次に、横野線座標を補正するため、横野線テーブル1105を参照する。横野線テーブル1105の全横野線を参照したかを判定する。全横野線を参照した場合には、ステップS29に進む。

【0090】(S28) 横野線の左側座標（開始座標） X_s 、 Y_s を、斜度 $\tan \theta$ に応じて補正する。補正後の座標X、Yは、前述の式(2)、(3)で示される。

【0091】そして、横野線の右側座標（終了座標） X_e 、 Y_e を、斜度 $\tan \theta$ に応じて補正する。補正後の座標X、Yは、前述の式(4)、(5)で示される。

【0092】そして、ステップS27に戻る。

【0093】(S29) 次に、縦野線座標を補正するため、縦野線テーブル1106を参照する。縦野線テーブル1106の全縦野線を参照したかを判定する。全縦野線を参照した場合には、終了する。

【0094】(S30) 縦野線の左側座標（開始座標） X_s 、 Y_s を、斜度 $\tan \theta$ に応じて補正する。補正後の座標X、Yは、前述の式(2)、(3)で示される。

【0095】そして、縦野線の右側座標（終了座標） X_e 、 Y_e を、斜度 $\tan \theta$ に応じて補正する。補正後の座標X、Yは、前述の式(4)、(5)で示される。

【0096】そして、ステップS29に戻る。

【0097】このようにして、横野線、縦野線、野線項

これらの各欄は、後の処理により利用される。

【0079】次に、斜度補正について、説明する。図30は、斜度算出処理フロー図である。

【0080】(S20) 横野線情報テーブル1105の各横野線を参照する。そして、全横野線を参照したかを判定する。

【0081】(S21) 参照した横野線の斜度を算出する。斜度 θ は、終了Y座標を Y_e 、開始Y座標を Y_s 、終了X座標を X_e 、開始X座標を X_s とすると、次の式(1)で計算される。

【0082】

(1)

野線項目を参照する。野線項目テーブル1104の全野線項目を参照したかを判定する。全野線項目を参照した場合には、ステップS27に進む。

【0086】(S26) 野線項目の左上座標（開始座標） X_s 、 Y_s を、斜度 $\tan \theta$ に応じて補正する。補正後の座標X、Yは、次式(2)、(3)で示される。

【0087】

(2)

(3)

Yは、次式(4)、(5)で示される。

【0088】

(3)

(4)

目（最小矩形）を抽出する。そして、帳票の斜行による座標の誤差を補正するため、横野線座標から斜度を計算する。斜度は平均値のため、正確に斜度を求めることができる。更に、斜度により、野線項目、横野線、縦野線の座標を補正する。

【0098】(b) グループ分類処理の説明

図32は、グループ処理フロー図、図33は、グループ処理後のテーブル更新の説明図である。

【0099】(S31) 野線項目テーブル1104から野線項目を参照する。全野線項目を参照したかを判定する。全野線項目を参照した場合には、終了する。

【0100】(S32) 野線項目テーブル1104から他の野線項目を抽出する。全部の他野線項目を参照したかを判定する。全野線項目を参照した場合には、ステップS35に進む。

【0101】(S33) 全野線項目が参照済でない場合には、抽出した2つの野線項目が接続関係にあるかを判定する。図8で説明したように、2つの野線項目の位置関係が、接続関係にあるかを判定する。接続関係にない場合は、ステップS32に戻る。

【0102】(S34) 2つの野線項目が接続関係にある場合には、1の野線項目が属するグループ情報テーブル1100に、他の野線項目を反映する。図33に示すように、グループ情報テーブル1100は、グループ番

号GN、開始座標SC、終了座標EC、野線項目数RIN、ブロック数BLN、論理行数LR、先頭野線項目アドレスHLA、先頭論理行情報アドレスHRA、先頭ブロックアドレスHBAをからなる。グループ情報テーブル1100のグループ矩形座標SC、ECと、野線項目数RINを更新する。更に、野線項目テーブル1104に、グループ番号を設定する。そして、ステップS32に戻る。

【0103】(S35) 全野線項目を参照したと判定すると、1の接続野線に、接続する野線項目はあるかを判定する。接続する野線項目がある時は、ステップS31に戻る。接続する野線項目がない場合には、グループ情報テーブル1100に、新規にグループ情報を作成する。そして、ステップS31に戻る。

【0104】図2の帳票例では、図33のグループ情報テーブル1100に示すように、2つのグループに、分類される。

【0105】このようにして、野線項目を表部とそれ以外を分離するため、各野線項目をグループに分類する。

【0106】(c) 構成要素分類処理の説明

図34及び図35は、構成要素分類処理フロー図、図36及び図37は、分類処理後のテーブル説明図である。

【0107】(S40) グループ情報テーブル1100からグループ情報を抽出する。全グループ情報を参照したかを判断する。全グループ情報を参照したと判定すると、終了する。

【0108】(S41) 横野線テーブル1105から、そのグループの横野線を抽出する。そのグループの横野線を全て参照したかを判定する。そのグループの横野線を全て参照したと判定すると、ステップS43に進む。

【0109】(S42) 横野線を比較して、そのグループの左右に達する横野線で、最小のY座標と最大のY座標を持つ横野線を抽出する。ステップS41に戻る。

【0110】(S43) 縦野線テーブル1106から、縦野線情報を抽出する。縦野線を全て参照したかを判定する。縦野線を全て参照したと判定すると、ステップS45(図35)に進む。

【0111】(S44) 最小のY座標と最大のY座標を持つ横野線の左右端と交点を持つ縦野線を2本抽出する。ステップS43に戻る。

【0112】(S45) 野線項目テーブル1104から、グループを構成する野線項目を参照する。グループを構成する全野線項目を参照したかを判定する。そのグループの全野線項目を参照した場合には、図34のステップS40に戻る。

【0113】(S46) ステップS42とステップS44により、そのグループの最大矩形を構成する2本の横野線と2本の縦野線が抽出された。そこで、野線項目が、この最大矩形の中にあるかを判定する。即ち、野線項目が、抽出された横野線と縦野線で囲まれる範囲(外

接矩形)内に位置するかを判定する。

【0114】(S47) 野線項目が、範囲内にある場合は、野線項目を当該グループに所属させる。更に、グループ情報と当野線項目の開始座標が一致する場合は、グループ情報の先頭野線項目アドレスHLAを、当野線項目に設定する。そして、ステップS45に戻る。

【0115】(S48) 当該野線項目が、範囲内にはない場合には、当該野線項目の開始Y座標と、外接矩形開始Y座標とを比較する。野線項目の開始Y座標が、外接矩形の開始Y座標より大きい場合には、その野線項目は、上凸グループに分類する。尚、上凸グループが作成済なら、上凸グループの野線項目と接続しているかを判定する。当該野線項目が、上凸グループの野線項目と接続している場合には、当該野線項目をそのグループに追加する。

【0116】下凸グループも同様であり、野線項目の開始Y座標が、外接矩形の開始Y座標より小さい場合には、その野線項目は、下凸グループに分類する。尚、下凸グループが作成済なら、下凸グループの野線項目と接続しているかを判定する。当該野線項目が、下凸グループの野線項目と接続している場合には、当該野線項目をそのグループに追加する。そして、ステップS45に戻る。

【0117】このようにして、各グループを、主表部と、上凸部、下凸部の表の構成要素に分類する。図2の帳票例では、2グループに分類された帳票が、4グループに分類される。即ち、図37のグループ情報テーブル1100に示すように、グループは、4つとなる。そして、図36の野線項目テーブル1104に示すように、野線項目も4つのグループに分類される。即ち、グループ番号は4つとなる。

【0118】(d) 論理行分類処理の説明

図38は、論理行分類処理フロー図、図39は、図38の入れ子解析処理フロー図、図39及び図40は、論理行抽出後のテーブル説明図である。

【0119】(S50) グループ情報テーブル1100からグループ情報を抽出する。全部のグループ情報を参照したかを判定する。全てのグループ情報を参照すると、終了する。

【0120】(S51) 横野線テーブル1105から、そのグループの左右端に交点を持つ横野線を2本ずつ抽出する。全横野線を抽出したかを判定する。そのグループの全横野線を抽出した時は、ステップS50に戻る。

【0121】(S52) 全横野線を参照していない場合には、野線項目テーブル1104からそのグループを構成し、2本の横野線に位置する野線項目を抽出する。全野線項目を抽出したかを判定する。そのグループの全野線項目を参照した時は、ステップS51に戻る。

【0122】(S53) 2本の横野線の内、下に位置する横野線座標が野線項目の下Y座標に一致するかを判定

する。

【0123】(S54)一致しない場合には、図12(A)及び図12(B)に示したように、入れ子構造の可能性はある。このため、後述する図39に示す入れ子構造の解析処理を行う。

【0124】(S55)論理行に分類するため、論理行テーブル1102を作成する。論理行テーブル作成済の場合には、当該野線項目を追加する。論理行テーブルが作成されていない場合には、新規に論理行テーブルを作成する。

【0125】このようにして、各グループにおいて、野線項目を、論理行に分類する。

【0126】次に、図39により、入れ子構造解析処理について、説明する。

【0127】(S56)野線項目テーブルから野線項目を2つずつ抽出する。例えば、1回目は、1番目と2番目、2回目は、2番目と3番目とする。

【0128】(S57)野線項目の開始Y座標と、論理行の上側横野線のY座標が一致し、野線項目の終了Y座標か、論理行の下側横野線のY座標が一致するかを判定する。一致する場合には、終了する。

【0129】(S58)一致しない場合には、2つの野線項目の開始X座標を比較する。2つの野線項目の開始X座標が一致する場合には、ステップS59に進む。2つの野線項目の開始X座標が一致する場合には、図12(A)に示すように、2つの野線項目に親子関係を設定する。そして、ステップS56に戻る。

【0130】(S59)X座標が一致しない場合には、2つの野線項目の開始Y座標を比較する。2つの野線項目の開始Y座標が一致しない場合には、ステップS56に戻る。2つの野線項目の開始Y座標が一致する場合には、図12(A)に示すように、2つの野線項目に兄弟関係を設定する。そして、ステップS56に戻る。

【0131】図41に示すように、論理行テーブル1102が作成される。論理行テーブルは、各論理行の番号LN、開始座標SC、終了座標EC、グループ番号GN、ブロック番号BN、野線項目数RIN、論理列数LL、上側横野線番号ULN、下側横野線番号LLN、左側縦野線番号ERN、右側縦野線番号RRN、先頭野線項目アドレスHLAからなる。図2の帳票の例では、図41に示すように、論理行テーブル1102が作成される。

【0132】更に、グループ情報テーブル1100は、論理行数LRと先頭論理行アドレスHLAが追加される。図40に示すように、野線項目テーブル1104の入れ子情報INは、入れ子フラグINF、階層番号KN、子野線項目番号CRN、野線項目間リンク番号RLNからなる。入れ子フラグINFは、野線項目が入れ子構造を構成する場合に、「ON」となる。図2の帳票例では、野線項目10、11、12が入れ子構造を構成す

るため、「ON」である。

【0133】階層番号KNは、入れ子構造の階層の深さ(親子関係数)を示す。図2の帳票例では、親となる野線項目10は、「0」、子となる野線項目11、12は、「1」を設定する。子野線項目数CRNは、当野線項目が親子の関係の場合には、親の野線項目に、子となる野線項目番号を設定する。図2の帳票例では、親の野線項目10に、子の野線項目番号11を設定する。

【0134】野線項目間リンク番号RLNは、野線項目が兄弟関係にある場合に、兄弟関係にある野線項目の通し番号を示す。図2の帳票例では、野線項目11、12が、兄弟関係にあるため、野線項目11に「0」を、野線項目12に「1」を設定する。

【0135】このようにして、各野線項目を論理行に分類する。この論理行は、同じ構造をしている行を見つけるのに役立つ。

【0136】次に、図3の帳票例2のように、表が分離されている場合の表の統合処理について、説明する。

【0137】図42は、表の統合処理フロー図、図43及び図44は、帳票例2のグループ統合前のテーブル説明図、図45及び図46は、帳票例2のグループ統合後のテーブル説明図である。

【0138】(S60)グループ情報テーブル1100から、同じ開始Y座標と終了Y座標を持つ2つのグループを抽出する。

【0139】(S61)2つのグループを構成する論理行情報を1つずつ抽出する。全部の論理行を参照したかを判定する。全部の論理行を参照した場合には、ステップS63に進む。

【0140】(S62)抽出した論理行情報から、開始Y座標と終了Y座標が一致する論理行数をカウントする。そして、ステップS61に戻る。

【0141】(S63)一致した論理行数の全論理行数に対する割合を計算する。そして、その割合と所定の閾値とを比較する。その割合が、閾値以下の場合には、2つのグループは別の表と判断して、終了する。

【0142】(S64)その割合が、閾値以上の場合には、2つのグループは、1つの表であると判定する。そして、2つのグループに、異なる構造の論理行はあるかを判定する。異なる構造の論理行がある場合には、図47で説明する野線作成処理を行う。

【0143】(S65)2つのグループを1つのグループに統合する。そして、終了する。

【0144】図3の帳票例で説明する。図3の帳票例では、野線項目テーブル1104は、図43のように設定される。論理行テーブル1102及びグループ情報テーブル1100は、図44に設定される。即ち、グループは、4つに分類される。前述した統合処理により、グループ2と4が統合される。即ち、図45に示すように、論理行テーブル1102において、グループ4の論理行

は削除される。グループテーブル1100においても、グループ4は削除される。同様に、論理行テーブル1102、グループ情報テーブル1100において、グループ2の罫線項目数に、グループ4の罫線項目数を加算される。罫線項目テーブル104において、グループ4の罫線項目のグループ番号は、グループ2に変更される。更に、横罫線テーブル1105において、グループ2の論理行を構成する横罫線の終了X座標を、グループ4の終了X座標まで延長する。

【0145】このようにして、グループの論理行構造が同じか、類似しないかを判定して、表の統合を行う。

【0146】次に、図42の罫線作成処理について、説明する。図47は、罫線作成処理フロー図、図48は、罫線作成の説明図、図49及び図50は、罫線作成後のテーブル説明図である。

【0147】(S66) 2つのグループから論理行情報を抽出する。全論理行を参照したかを判定する。全論理行を参照した場合には、終了する。

【0148】(S67) 2つの論理行の高さは、一致するかを判定する。一致する場合は、罫線作成処理不要のため、ステップS66に戻る。一致していない場合には、次の行の論理行情報を抽出する。そして、1のグループの次の行Bと行Aの高さとの和が、他のグループの行Cの高さとが一致するかを判定する。図48に示すように、グループ2の論理行A、Bと、グループ4の論理行Cが対応するものとする。この場合に、グループ4の論理行Cにおいて、掠れにより罫線が消失している。この場合に、グループ2の次の行Bと行Aの高さとの和が、グループ4の行Cの高さとが一致する。一致しない場合は、ステップS66に戻る。

【0149】(S68) 一致する場合には、行Cの中で、行Aと行Bの境界線と同じY座標を持つ横罫線を生成する。そして、1つの行を2つの行に分割する。そして、ステップS66に戻る。

【0150】このようにして、掠れ等で消失した罫線を、表構造から検出する。そして、罫線を作成する。図49に示すように、罫線項目情報テーブル1104において、罫線項目158と罫線項目159の下側横罫線番号を新規に作成した横罫線番号に変更する。終了Y座標を横罫線の座標に変更する。罫線項目160と罫線項目161を新たに作成する。図50の論理行テーブル1102において、行Cを上下に2分割する。下側論理行を論理行テーブルに追加する。グループテーブル1100において、グループ4の論理行数を+1する。

【0151】(e) ブロック分類処理の説明

図51は、ブロック分類処理フロー図、図52及び図53はブロック分類後のテーブル説明図である。

【0152】(S70) グループ情報テーブル1100から、グループ情報を抽出する。全てのグループ情報を抽出したかを判定する。全てのグループ情報を抽出した

場合には、終了する。

【0153】(S71) そのグループを構成する論理行情報を2行ずつ抽出する。例えば、1回目は1行目と2行目、2回目は2行目と3行目を抽出する。全ての論理行を抽出したかを判定する。そのグループの全ての論理行を抽出した場合には、ステップS70に戻る。

【0154】(S72) 抽出した2つの論理行の構造を比較する。この比較は、次の項目を比較する。構成する罫線項目数は同じか。論理行の中で、罫線項目の配置は一致するか。論理行の高さ、幅は同じか。

【0155】(S73) 論理行構造が一致しない場合には、両論理行は別のブロックを構成すると判断する。このため、両論理行をリンクしないブロック情報を作成する。そして、ステップS70に戻る。

【0156】(S74) 論理行構造が一致する場合には、両論理行は同一のブロックを構成すると判断する。このため、両論理行をリンクするブロック情報を作成する。そして、ステップS70に戻る。

【0157】このようにして、グループ単位に、行構造が同一である論理行を1つのブロックに分類する。図53に示すように、ブロック情報テーブル1101は、ブロック番号BN、開始座標SC、終了座標EC、グループ番号GN、罫線項目数RIN、論理行数LN、論理列数RO、先頭論理行アドレスHRA、先頭論理列アドレスHRA、上側横罫線番号ULN、下側横罫線番号LLN、左側縦罫線番号LVN、右側縦罫線番号RVNからなる。

【0158】ブロック情報テーブル1101に、各ブロックの開始座標SC、終了座標EC、グループ番号GN、罫線項目数RIN、論理行数LN、先頭論理行アドレスHRA、上側横罫線番号ULN、下側横罫線番号LLN、左側縦罫線番号LVN、右側縦罫線番号RVNを設定する。

【0159】そして、図52に示すように、罫線項目テーブル1104に、各罫線項目のブロック番号BNを設定する。論理行テーブル1102に、各論理行のブロック番号BNを設定する。図53に示すように、グループテーブル1100に、各グループのブロック数BLN、先頭ブロック番号アドレスを設定する。

【0160】次に、ブロック分類において、異なるブロックに分類されたブロックを統合する処理について、説明する。図54は、罫線回復処理の説明図、図55は、罫線回復処理フロー図、図56及び図57は、罫線回復前のテーブル説明図、図58乃至図60は、罫線回復後のテーブル説明図である。

【0161】図54に示すように、図2の帳票例1において、罫線項目63と罫線項目64の部分が掠れにより、罫線が消失しているものとする。この部分は、上下の論理行と構造が異なるため、ブロックBに分類される。又、ブロックBは、論理行が1行の構造であるた

め、野線項目の抽出順序は、他の行と異なっている。以下、図55により、処理を説明する。

【0162】(S75) グループ情報テーブル1100からグループ情報を抽出する。全てのグループ情報をちゅしゅつしたかを判定する。全てのグループ情報を抽出した場合には、処理を終了する。

【0163】(S76) ブロック情報テーブル1101からそのグループのブロック情報を抽出する。全ブロック情報を抽出したかを調べる。全ブロックを抽出した場合には、ステップS79に進む。

【0164】(S77) ブロック情報テーブル1101から次のブロック情報を抽出する。全ブロック情報を抽出したかを調べる。全ブロックを抽出した場合には、ステップS76に戻る。

【0165】(S78) 2つのブロック情報を構成する論理行の構造を比較する。即ち、ブロックを構成する論理行を構成する野線項目の数は同じかを判定する。論理行の中で野線項目の配置は一致するかを判定する。論理行の高さ、幅は同じかを判定する。一致する場合には、ブロック数をカウントする。そして、一致するブロック情報同志に、マーキングする。そして、ステップS77に戻る。

【0166】(S79) ブロック情報テーブル1101から一致数が最も多いブロックを基準ブロックとし、一致数が最も多くないブロックを抽出する。図54で説明すると、基準ブロックは、ブロックAが抽出され、一致数が最も多くないブロックとして、ブロックBが抽出される。全ブロック情報を抽出したかを調べる。全ブロックを抽出した場合には、ステップS75に戻る。

【0167】(S80) 基準ブロックAとブロックBの論理行での野線項目の配置を比較する。即ち、野線項目の座標を論理行開始X座標、Y座標からの相対座標に変換して、比較する。そして、野線項目の配置が異なった場合に、ブロックAの野線項目の高さh1とブロックBの野線項目の高さh2とを比較する。この比較において、 $h1 < h2$ である場合には、野線項目の野線の消失と判断する。このため、ブロックBの野線項目を、ブロックAの野線項目の高さと一致するように、仮分割して、ブロックBについて、論理行情報を仮作成する。そして、ブロックAの論理行と仮作成したブロックBの論理行が一致する時に、グループ情報、ブロックBのブロック情報と論理行情報、野線項目情報と横野線情報を編集する。そして、ステップS79に戻る。

【0168】これを、前述の図54の例で説明する。野線回復前のテーブルは、図56及び図57の如くである。即ち、図56の野線項目テーブル1104に示すように、野線項目59～野線項目66は、ブロック3に分類されている。野線項目47、48は、ブロック2に分類されている。

【0169】野線回復後は、図58に示すように、野線

項目テーブル1104において、野線項目63と64の終了座標を、横野線12の座標値に変更する。野線項目177と178を新規に作成する。上側が横野線11、下側が横野線12の野線項目と、上側が横野線11、下側が横野線13の野線項目とのブロック番号を、ブロック2に変更する。その野線項目の論理行番号をブロック2内の通し番号に変更する。下側が横野線13の野線項目の下側横野線番号を「12」に変更する。同一グループ内の野線番号が、論理行順になるように、テーブル1104をソートする。同一論理行番号の野線項目をX座標順になるようにテーブルをソートする。グループ3、ブロック4以降の野線項目のブロック番号をブロック2に変更し、論理行番号をブロック2内の通し番号に変更する。

【0170】図59の論理行テーブル1102において、グループ3、ブロック3の論理行情報を横野線11、12に囲まれる範囲に変更する。グループ3、ブロック3の論理行情報のブロック番号を、ブロック2に変更する。論理行番号をブロック2内の通し番号(4)に変更する。グループ3、ブロック2の野線項目数を「14」に、下側横野線番号を「12」に変更する。グループ3、ブロック2の論理行番号を新規に作成する。グループ3、ブロック4以降の論理行情報のブロック番号を、ブロック2に変更する。論理行番号をブロック2内の通し番号に変更する。

【0171】図59のブロック情報テーブル1101において、グループ3、ブロック2の野線項目数を「140」に、論理行数を「10」に、下側横野線番号を「19」に変更する。グループ3のブロック3とブロック4を、テーブルから削除する。

【0172】図60のグループ情報テーブル1100において、グループ2のブロック数を「2」に、野線項目数を「147」に変更する。横野線テーブル1105において、横野線12の終了Y座標を、横野線11と同じ座標値に変更する。横野線13を削除する。

【0173】このようにして、隣接する異なるブロック間において、論理行の構造を比較する。これにより、基準ブロックと構造の似ているブロックを探し、似ているブロックを基準ブロックの論理行構造になるように、野線を付加する。これにより、掠れた野線を回復する。

【0174】次に、帳票の野線項目に消し線が記入されて、異なるブロックに分類されたものを、同一のブロックに統合する処理について、説明する。図61は、論理行統合処理の説明図、図62は論理行統合処理フロー図、図63及び図64は、ろた行統合前のテーブル説明図、図65乃至図67は、論理行統合後のテーブル説明図である。

【0175】図61は、帳票の野線項目に消し線が記入された例を示し、野線項目59と野線項目73との間に、横野線が消し線として記入されている。元の帳票で

は、罫線項目59と罫線項目73とが合わせて1つの罫線項目であった。消し線により、1つの罫線項目が、2つに分割されている。2つに分割された行を1つの行に統合する。又、記入内容の訂正を行うため、引かれた訂正線により、罫線項目が部分的に分割された場合にも、統合処理により解析が可能となる。

【0176】図62の処理フローにより、説明する。

【0177】(S81)図55の罫線回復処理のステップS75～ステップS79の処理を実行して、基準ブロックと一致数が異なるブロックを抽出する。この参照中のブロックは、基準ブロックかを判定する。参照されたブロックが基準ブロックの場合には、ステップS79(図55)に戻る。

【0178】(S82)論理行テーブル1102から、参照ブロックの論理行情報を抽出する。参照ブロックの先頭論理行の高さに、その下の論理行の高さを1行ずつ加算する。加算した高さが、基準ブロックの高さに、一致するまで、論理行を参照する。一致しなければ、ステップS79(図55)に戻る。

【0179】(S83)一致した場合には、参照した論理行を構成する罫線項目の内、上下に接続する罫線項目を全て統合する。即ち、同じY座標を持ち、且つ共有するY座標がある罫線項目を統合する。更に、統合した罫線項目で構成される論理行を仮作成する。そして、その作成した論理行と基準ブロックの論理行の構造を比較する。論理行構造が一致する場合に、グループ情報、ブロック情報、論理行情報、罫線項目情報を編集する。そして、ステップS79(図55)に戻る。

【0180】このようにして、基準ブロックと、基準ブロックと構造の異なる他のブロックを抽出し、他のブロックの高さと基準ブロックの高さから、消し線の存在を検出する。そして、消し線の存在を検出すると、上下に位置する罫線項目を統合し、論理行を仮作成する。仮作成した論理行の構造と基準ブロックの論理行の構造を比較して、確認する。このため、消し線、訂正線を除去することができる。

【0181】図61のように、消し線が引かれた場合には、罫線項目テーブル1104は、図63のように、作成される。論理行情報テーブル1102、ブロック情報テーブル1101、グループ情報テーブル1100の各々は、図64に示すように、作成される。

【0182】図61の消し線が除去され、ブロックA、B、Cが、ブロックAに統合されると、図65に示すように、罫線項目テーブル1104では、罫線項目59から罫線項目72までの終了座標を横罫線13の座標値に、下側横罫線番号を「13」に変更する。罫線項目59から罫線項目72までのブロック番号を、ブロック2に変更する。罫線項目59から罫線項目72までの論理行番号を、ブロック2内での通し番号に変更する。罫線項目73から罫線項目86までを削除する。罫線項目1

01以降のブロック番号をブロック2に、論理行番号をブロック2内の通し番号に変更する。

【0183】図66に示すように、論理行情報テーブル1102において、グループ3、ブロック3の論理行終了Y座標を、横罫線13のY座標に変更し、下側横罫線番号を「13」に変更する。グループ3、ブロック3のブロック番号を、ブロック2に変更し、論理行番号を、ブロック2内の通し番号(4)に変更する。グループ3、ブロック4以降のブロック番号をブロック2に変更し、論理行番号をブロック2内の通し番号に変更する。

【0184】図66に示すように、ブロック情報テーブル1101において、グループ3、ブロック2の罫線項目数を「140」に、論理行数を「10」に、下側横罫線番号を「19」に変更する。グループ3のブロック3とブロック4とをテーブルから削除する。

【0185】図67に示すように、グループ情報テーブル1100において、グループ2の罫線項目数を「147」に、ブロック数を「2」に、論理行数を「11」に変更する。横罫線テーブル1105において、横罫線12を削除する。

【0186】(f)論理列判定処理の説明

図68は、論理列判定処理フロー図、図69乃至図71は、論理列判定後のテーブル説明図である。論理列の判定処理について、図68により説明する。

【0187】(S90)グループ情報テーブル1100から、グループ情報を抽出する。全てのグループ情報を抽出したかを判定する。全てのグループ情報を抽出した場合には、終了する。

【0188】(S91)ブロック情報テーブル1101から、抽出したグループの2つのブロックのブロック情報を抽出する。抽出したグループの全てのブロック情報を抽出したかを判定する。全てのブロック情報を抽出した場合には、ステップS90に戻る。

【0189】(S92)罫線項目テーブル1104から、抽出した各ブロックの先頭の論理行を構成する罫線項目情報を順次抽出する。抽出した論理行の全ての罫線項目情報を参照したかを判定する。全ての罫線項目情報を参照した場合には、後述するステップS97で、入れ子情報を設定したかを判定する。入れ子情報を設定していない場合には、抽出した2つのブロック間の論理列が終了するので、ステップS91に戻る。入れ子情報を設定した場合には、グループ情報、ブロック情報、論理行情報を更新する。更に、参照中ブロックの2行以下の論理行についても、1行目と同様に、各情報を更新する。そして、ステップS91に戻る。

【0190】(S93)2つの論理行の罫線項目が、同じ幅かを判定する。同じ幅でない場合には、幅の狭い論理行側で、次の罫線項目を抽出する。そして、幅の狭い論理行側で、現罫線項目と次の罫線項目との幅を加算する。幅の広い論理行の罫線項目と、その加算値を比較し

て、同じ幅となるかを判定する。同じ幅となるまで、野線項目の抽出を繰り返す。

【0191】(S94)このようにして、第1のブロックの先頭論理行の野線項目と、第2のブロックの先頭論理行の野線項目との間の列方向の対応関係が判明した。

【0192】次に、抽出した野線項目数をカウントする。

【0193】(S95)入れ子構造を検出するため、抽出した野線項目の数が「1」かを判定する。

【0194】(S96)野線項目数が「1」である場合には、入れ子構造でない。このため、抽出した野線項目に、左側からの通し番号を、論理列番号として設定する。

【0195】(S97)野線項目数が「1」でない場合には、入れ子構造である。このため、入れ子構造の野線項目に、同じ通し番号を、論理列番号として設定する。そして、入れ子構造を構成する野線項目に、入れ子情報を設定する。

【0196】(S98)論理列情報を作成し、同じ論理列内の野線項目間にリンクを設定する。この時、同じグループで異なるブロック、論理行間で同じ列番号が設定されたものの間に、リンクを設定する。そして、ステップS92に戻る。

【0197】このようにして、上下に接続するブロックの間で、同じ列構造を持つかを判定する。そして、同じ列構造を持つ場合、それらブロックを論理列の単位で統合する。

【0198】図2の帳票例で説明する。図69に示すように、野線項目情報テーブル1104では、野線項目の列番号欄ROに、論理列番号が設定される。例えば、ブロック1の野線項目10、11、12には、列番号が「1」と設定される。異なるブロック2において、この野線番号と同じ幅の野線項目は、「17」と「18」である。従って、ブロック2の野線項目17、18には、同じ列番号「1」が設定される。

【0199】次に、入れ子構造を構成する野線項目に入れ子情報INを設定する。図69では、野線項目17、18に、入れ子情報を設定する。同様に、同じ論理列を構成する野線項目(例えば、野線項目31、32)にも、入れ子情報を設定する。更に、同じ論理列を構成する野線項目間に、アドレスリンクHAAを設定する。

【0200】この場合に、入れ子情報INは、入れ子フラグINF、階層番号KN、子野線項目番号CRN、野線項目間リンク通し番号RLNからなる。野線項目17、18の例では、野線項目17、18が入れ子構造のため、入れ子フラグINFがオンとなる。2つの野線項目17、18間に親子関係がないため、階層番号KNに、「0」を設定する。2つの野線項目17、18間に親子関係がないため、子野線項目番号CRNに、「0」を設定する。野線項目17、18が兄弟関係のため、野

線項目17の野線項目間リンク番号RLNに、「0」を、野線項目18の野線項目間リンク番号RLNに、「1」を設定する。

【0201】図70に示すように、論理行情報テーブル1102において、論理行2の論理列数LLを、「5」に変更する。図70に示す論理列情報テーブル1103は、新規に作成する。この論理列情報テーブル1103は、論理列番号RO、開始座標SC、終了座標EC、グループ番号GN、ブロック番号BN、野線項目数RIN、上側横野線番号ULN、下側横野線番号LLN、左側縦野線番号ELN、右側縦野線番号RRN、先頭野線項目アドレスHLAからなる。前述の列方向の関係に応じて、これらを作成する。

【0202】更に、図71に示すように、ブロック情報テーブル1101は、グループ3、ブロック2の論理列数が、「5」に変更される。そして、各ブロックの先頭論理列アドレスHAAが設定される。

【0203】このようにして、各グループにおいて、ブロック間の野線項目の列方向の関係が判定される。そして、これに従い、列方向の関係を記述するテーブルが更新される。

【0204】次に、図20(A)及び図20(B)で説明した1桁及び3桁野線の入れ子解析処理について、説明する。図72は、1桁、3桁野線の入れ子解析処理フロー図、図73は入れ子解析後のテーブル説明図である。

【0205】(S100)論理行テーブル1102から論理行情報を抽出する。全論理行情報を抽出したかを判定する。全論理行情報を抽出した場合には、終了する。

【0206】(S101)論理行情報からリンクする野線項目を抽出する。全野線項目を抽出したかを判定する。全野線項目を抽出した場合には、ステップS100に戻る。

【0207】(S102)抽出した野線項目のサイズが、1桁野線項目のサイズに該当するかを判定する。該当する場合には、ステップS104に進む。該当しない場合には、ステップS103に進む。

【0208】(S103)抽出した野線項目のサイズが、3桁野線項目のサイズに該当するかを判定する。該当する場合には、ステップS104に進む。該当しない場合には、ステップS101に戻る。

【0209】(S104)参照中の論理行の中で、次の野線項目を抽出する。現野線項目と次の野線項目との大きさを比較する。誤差範囲内で一致するかを判定する。一致する場合には、ステップS105に進む。一致しない場合には、ステップS101に戻る。

【0210】(S105)現野線番号のリンク先野線番号に、次の野線番号を設定する。そして、ステップS101に戻る。

【0211】図73には、図2の帳票例における野線項

目テーブル1104の更新後の状態を示す。図2において、1桁罫線項目は、罫線項目22~30、36~44等である。上述の処理により、これらの罫線項目22~30が、1桁罫線項目と判定される。そして、罫線項目テーブル1104において、網掛け状態にして示すように、1桁罫線項目22~30のリンク先罫線項目番号ANに、リンクする罫線番号が設定される。尚、前述の入れ子処理により、罫線番号22~30は、入れ子構造と判定され、入れ子情報INが設定されている。

【0212】このようにして、予め1桁罫線項目又は3桁罫線項目の縦サイズと横サイズを定義しておく。これらサイズと一致する罫線項目が任意数並んだ場合には、これらを入れ子構造として、論理的に1つの罫線項目としてまとめる。このような結果を利用して、文字認識処理では、連続する罫線項目を1つの文字列として扱うことができる。

【0213】次に、図21(A)及び図21(B)に示した特殊な構造を持つ矩形の入れ子構造を解析する処理について、説明する。図74は、主表以外の入れ子解析処理フロー図、図75及び図76は、入れ子解析後のテーブル説明図である。

【0214】(S110) グループ情報テーブル1100から、グループ情報を抽出する。全グループ情報を抽出したかを判定する。全グループ情報を抽出した場合には、終了する。

【0215】(S111) 抽出したグループの構造を判定する。グループは、1ブロック且つ1行かを判定する。グループが、1ブロック且つ1行である場合には、ステップS114に進む。グループが、1ブロック且つ1行でない場合には、グループが1列かを判定する。グループが1列でない場合には、ステップS110に戻る。グループが1列である場合には、グループが複数ブロックの構造かを判定する。グループが複数ブロックの構造でない場合には、ステップS110に戻る。グループが複数ブロックの構造である場合には、ステップS112に進む。

【0216】(S112) そのグループを構成する罫線項目を抽出する。全罫線項目情報を参照したかを判定する。全罫線項目情報を参照した場合には、ステップS110に戻る。

【0217】(S113) そのグループにおいて、現参照罫線項目の次の罫線項目を順次抽出する。そして、現罫線項目と次の罫線項目の左側縦罫線番号を比較する。両罫線項目の左側縦罫線番号が一致する場合には、罫線項目間に入れ子情報(親子関係)を設定する。次に、現罫線項目と次の罫線項目の上側横罫線番号を比較する。両罫線項目の上側横罫線番号が一致する場合には、罫線項目間に入れ子情報(兄弟関係)を設定する。これにより、図21(B)に示す例の入れ子構造が検出される。そして、ステップS112に戻る。

【0218】(S114) グループが、1ブロック且つ1行である場合には、そのグループを構成する罫線項目を抽出する。全罫線項目情報を参照したかを判定する。全罫線項目情報を参照した場合には、ステップS110に戻る。

【0219】(S115) そのグループにおいて、現参照罫線項目の次の罫線項目を順次抽出する。そして、現罫線項目と次の罫線項目の上側横罫線番号を比較する。両罫線項目の上側横罫線番号が一致する場合には、罫線項目間に入れ子情報(親子関係)を設定する。次に、現罫線項目と次の罫線項目の左側縦罫線番号を比較する。両罫線項目の左側縦罫線番号が一致する場合には、罫線項目間に入れ子情報(兄弟関係)を設定する。これにより、図21(A)に示す例の入れ子構造が検出される。そして、ステップS114に戻る。

【0220】このようにして、帳票上の主表以外の箇所を、構造毎に入れ子構造解析処理を行う。図21(A)では、罫線項目1に対し、罫線項目2が親子の関係にある。罫線項目2に対し、罫線項目4が親子の関係にある。罫線項目3に対し、罫線項目5が親子の関係にある。罫線項目2に対し、罫線項目3が兄弟の関係にある。罫線項目4に対し、罫線項目5が兄弟の関係にある。図21(B)では、罫線項目1に対し、罫線項目2が親子の関係にある。罫線項目2に対し、罫線項目4が親子の関係にある。罫線項目3に対し、罫線項目5が親子の関係にある。罫線項目2に対し、罫線項目3が兄弟の関係にある。罫線項目4に対し、罫線項目5が兄弟の関係にある。

【0221】図2の帳票例において、各テーブルを説明する。図75に示すように、罫線項目テーブル1104では、罫線項目1~5に、入れ子フラグINFを設定される。罫線番号2、3に対しては、入れ子階層番号KNが「1」に設定され、罫線番号4、5に対しては、入れ子階層番号KNが「2」に設定される。子階層番号CRNは、罫線番号1に「2」が、設定され、罫線番号2に「4」が設定され、罫線番号3に「5」が設定される。罫線項目間リンク通し番号RLNは、罫線番号3と5に対し、「1」が設定される。

【0222】図75に示すように、論理行情報テーブル1102では、グループ1、ブロック1の論理行1に、論理列数「1」が設定される。図76に示すように、論理列情報テーブル1103では、論理列2、3が削除され、論理列番号1の罫線項目数が「5」に変更される。図76に示すように、ブロック情報テーブル1101では、グループ1、ブロック1の論理列数が「1」に変更される。

【0223】このようにして、主表以外の部分の入れ子構造を解析する。

【0224】(g) 属性付与処理の説明

図77は、属性処理フロー図、図78は、属性付与後の

テーブル説明図である。

【0225】(S120) グループ情報テーブル1100からグループ情報を抽出する。全グループ情報を抽出したかを判定する。全グループ情報を抽出した場合には、終了する。

【0226】(S121) そのグループは、複数ブロックで構成されるかを判定する。複数ブロックで構成されていない場合には、ステップS120に戻る。

【0227】(S122) グループ情報のブロック情報アドレスからブロック情報を抽出し、各ブロックにブロック属性BZを付与する。即ち、第1番目のブロックに、ヘッダブロック属性を付与し、第2番目のブロックに、ボディブロック属性を付与し、第3番目以降のブロックに、フッタブロック属性を付与する。そして、ステップS120に戻る。

【0228】このようにして、表構造部を構成するブロックに、属性を付与して、各ブロックに意味を持たせる。図78に示すように、ブロック情報テーブル1101において、グループ3の第1番目のブロック1に、ヘッダブロック属性「1」を付与し、第2番目のブロック2に、ボディブロック属性「2」を付与し、第3番目のブロック3に、フッタブロック属性「3」を付与する。

【0229】(h) 見出しフラグ設定処理の説明次に、見出しを構成する罫線項目に、見出しフラグを設定する見出しフラグ設定処理について、説明する。

【0230】図79は、フラグ設定処理フロー図、図80は、フラグ設定後のテーブル説明図である。

【0231】(S130) グループ情報テーブル1100から、グループ情報を抽出する。全てのグループ情報を抽出したかを判定する。全てのグループ情報を抽出した場合には、終了する。

【0232】(S131) グループ情報からアドレス指示されているブロック情報を抽出する。グループ内の全ブロック情報を抽出したかを判定する。グループ内の全ブロック情報を抽出した場合には、ステップS130に戻る。

【0233】(S132) 抽出したブロックは、ヘッダブロックかを判定する。ヘッダブロックである場合には、当該ブロックの論理行情報を1レコードずつ抽出して、見出しフラグをオンする。これにより、ヘッダブロックを構成する罫線項目全てに、見出しフラグを設定する。そして、ステップS131に戻る。

【0234】(S133) 抽出したブロックは、フッタブロックかを判定する。フッタブロックでない場合には、ステップS134に進む。フッタブロックである場合には、当該ブロックの論理行情報を1レコードずつ抽出する。論理行内の罫線項目情報を1レコードずつ抽出して、見出しフラグをオンする。これにより、フッタブロックを構成する罫線項目全てに、見出しフラグを設定

する。そして、ステップS131に戻る。

【0235】(S134) 抽出したブロックは、ボディブロックかを判定する。ボディブロックでない場合には、ステップS131に戻る。ボディブロックである場合には、当該ブロックの先頭論理行情報を1レコードずつ抽出する。抽出した先頭論理行内の罫線項目情報を1レコードずつ抽出して、見出しフラグをオンする。これにより、ボディブロックの先頭論理行を構成する罫線項目全てに、見出しフラグを設定する。そして、当該ブロックの先頭論理列情報を抽出する。抽出した先頭論理列内の罫線項目情報を1レコードずつ抽出して、見出しフラグをオンする。これにより、先頭論理列の罫線項目に、見出しフラグを設定する。そして、ステップS131に戻る。

【0236】この見出し部を抽出しておくこと、使用される文字を制限でき、その文字(文言)に特化した文字認識処理を行うことができる。従って、高精度に見出し部分の文言を認識することができる。図80に示すように、罫線項目テーブル1102において、図79のルールに従い、見出しと推定される罫線項目に、見出しフラグMFが設定される。

【0237】(i) 見出し部分の文字認識処理の説明次に、見出しの種類を決定するため、見出しフラグが設定された罫線項目内に存在する文字列の文字認識を行う。

【0238】図81は、見出し認識処理フロー図、図82は、見出し認識後のテーブル説明図である。

【0239】(S140) グループ情報テーブル1100から、グループ情報を抽出する。全てのグループ情報を抽出したかを判定する。全てのグループ情報を抽出した場合には、終了する。

【0240】(S141) グループ情報からアドレス指示されているブロック情報を抽出する。グループ内の全ブロック情報を抽出したかを判定する。グループ内の全ブロック情報を抽出した場合には、ステップS140に戻る。

【0241】(S142) ブロック情報はボディブロックかを判定する。ボディブロックである場合には、データ部分のため、ステップS141に戻る。

【0242】(S143) ボディブロックでない場合には、見出し部分の可能性がある。このため、ブロック情報からアドレス指示されている論理行情報を抽出する。全ての論理行情報を抽出したかを判定する。全ての論理行を抽出した場合には、ステップS141に戻る。

【0243】(S144) 論理行情報からアドレス指示されている罫線項目情報を抽出する。

【0244】(S145) そして、罫線項目情報の見出しフラグがオンかを判定する。見出しフラグがオンでない場合には、ステップS144に戻る。

【0245】(S146) 罫線項目内の文字列データ部

分のイメージを切り出す。そして、切り出したイメージを見出し文言辞書1003(図25参照)を使用して、文字認識する。図25に示したように、見出し文言辞書1003は、見出しとして存在する文言(振込先等)と、そのイメージとが登録されたものである。従って、一般の文字認識辞書を使用した場合よりも高い精度での文字認識が可能となる。

【0246】(S147)一致した辞書レコードの見出し名称を抽出する。見出しテーブル(図24参照)から見出し名称が一致するレコードを抽出する。即ち、認識結果から得られた見出し名称をキーとして、図24に示す見出しテーブルを参照する。図24に示したように、見出しテーブルには、見出し名称に対応する属性情報(文字認識カテゴリ、文字フォント等)が登録されている。そして、その見出し名称に一致するレコード(属性情報)を抽出する。そして、その野線項目情報に、属性(文字認識カテゴリCC、文字フォントCF)を設定する。そして、ステップS144に戻る。

【0247】このようにして、見出し部分を、限定された文字認識辞書を用いて、文字認識する。図82に示すように、図2に示す帳票例1の場合に、野線項目テーブル1104の見出しフラグがオンである野線項目に、文字認識カテゴリCC、文字フォントCFが設定される。

【0248】(j)データ部分の属性設定処理の説明次に、データ部分の属性を決定するため、見出しフラグが設定されていない野線項目の属性を決定する。

【0249】図83は、属性設定処理フロー図、図84は、属性設定後のテーブル説明図である。

【0250】(S150)グループ情報テーブル1100から、グループ情報を抽出する。全てのグループ情報を抽出したかを判定する。全てのグループ情報を抽出した場合には、終了する。

【0251】(S151)グループ情報からアドレス指示されているブロック情報を抽出する。グループ内の全ブロック情報を抽出したかを判定する。グループ内の全ブロック情報を抽出した場合には、ステップS150に戻る。

【0252】(S152)ブロック情報からアドレス指示されている論理列情報を抽出する。全ての論理列情報を抽出したかを判定する。全ての論理列を抽出した場合には、ステップS151に戻る。

【0253】(S153)論理列情報からアドレス指示されている野線項目情報を抽出する。そして、野線項目情報の見出しフラグがオンかを判定する。見出しフラグがオンでない場合には、ステップS155に進む。

【0254】(S154)見出しフラグがオンされている場合には、その野線項目の文字認識カテゴリ、文字フォントを保存する。そして、ステップS153に戻る。

【0255】(S155)見出しフラグがオンでない場合には、データ部分の野線項目である。野線項目情報

は、入れ子情報かを判定する。入れ子情報である場合には、ステップS156に進む。入れ子情報でない場合には、ステップS154で保存した文字認識カテゴリ、文字フォントを参照中の野線項目に設定する。これにより、データ部分に、同一論理列の見出し部分の属性が設定される。そして、ステップS153に戻る。

【0256】(S156)入れ子情報である場合には、参照中の野線項目と同一レベル(階層番号リンク通し番号)の野線項目から、ステップS154で保存した文字認識カテゴリ、文字フォントを参照中の野線項目に設定する。そして、ステップS153に戻る。

【0257】このようにして、論理列を構成する野線項目のうち、見出し以外の野線項目に対して、文字認識カテゴリ、文字フォント等の認識特性を、見出しの野線項目から複写する。図84に示すように、図2の帳票例の場合には、野線項目情報テーブル1104のデータ部分の野線項目3、5、7等に、同一論理列の野線項目の文字認識カテゴリCC、文字フォントCFが設定される。このため、論理列情報により、データ部分のデータ名属性が決定される。

【0258】(k)データ部分の文字認識処理の説明次に、データ部分の文字を認識するため、見出しフラグが設定されていない野線項目の文字認識を行う。

【0259】図85は、文字認識処理フロー図、図86は、文字認識結果の説明図である。

【0260】(S160)図86に示す文字認識結果通知領域管理部1200をメモリに設定する。管理部1200は、認識結果見出し数と、認識結果見出し領域アドレスとからなる。

【0261】(S161)グループ情報テーブル1100から、グループ情報を抽出する。全てのグループ情報を抽出したかを判定する。全てのグループ情報を抽出した場合には、終了する。

【0262】(S162)グループ情報からアドレス指示されているブロック情報を抽出する。グループ内の全ブロック情報を抽出したかを判定する。グループ内の全ブロック情報を抽出した場合には、ステップS161に戻る。

【0263】(S163)ブロック情報からアドレス指示されている論理列情報を抽出する。全ての論理列情報を抽出したかを判定する。全ての論理列を抽出した場合には、ステップS162に戻る。

【0264】(S164)論理列情報からアドレス指示されている野線項目情報を抽出する。

【0265】(S165)野線項目情報の見出しフラグがオンかを判定する。見出しフラグがオンである場合には、ステップS164に戻る。

【0266】(S166)見出しフラグがオンでない場合には、その野線項目内に存在する文字列を、一般の文字認識辞書1004(図1参照)を使用して、文字認識

する。この時、文字認識辞書 1004 は、漢字、数字等の各認識カテゴリに対し、別々に設けられている。従って、前述したデータ部分の認識属性に従い、一般辞書から認識辞書を選択して、文字認識する。このため、データ部分も高い精度で文字認識できる。

【0267】更に、図 86 に示す認識結果見出し領域 1201 を設定する。領域 1201 は、見出し名称と、認識結果データ数、認識結果レコード長、認識結果データアドレスから成る。そして、認識結果データ領域 1202 を設定する。即ち、認識結果データアドレスの示す領域に、認識結果を設定する。

【0268】このようにして、得られた認識結果を認識結果通知領域に設定する。これにより、データ部分のデータ名称、データ数、文字認識結果が得られる。

【0269】このようにして、各罫線項目の行と列の関係を判定し、見出し部とそれに対応するデータ部とに分類する。そして、見出し部分を見出し認識辞書を用いて文字認識して、見出し名称を得る。論理列の構造の関係から見出し部分に対応したデータ部分のデータ名称が判明する。このため、帳票の各項目のフォーマットを予め定義しておく必要がない。

【0270】更に、見出し名称から認識属性を見出しテーブルにより得る。これに従い、データ部分を文字認識する。このため、帳票の表構造を解析することができ、データ部分の文字認識結果を精度を高めることができる。

【0271】上述の実施の形態では、各帳票の記入又は印刷文字として、漢字、英字、数字を例に説明したが、数字又は英字のみであって良い。又、罫線項目の属性として、そのデータ名称、文字認識カテゴリ、文字フォントを検出しているが、データ名称のみであっても良い。

【0272】以上、本発明の実施の形態により説明したが、本発明の主旨の範囲内で種々の変形が可能であり、これらを本発明の範囲から排除するものではない。

【0273】

【発明の効果】以上説明したように、本発明によれば、次の効果を奏する。

【0274】(1) 帳票の罫線項目の並びを検出して、帳票の表構造を自動解析するため、帳票のフォーマットを登録する手間を省くことができる。

【0275】(2) 帳票のフォーマットが自動登録されるので、フォーマットを変更しても、再登録の手間を省くことができる。

【図面の簡単な説明】

【図 1】本発明の一実施の形態のブロック図である。

【図 2】本発明の一実施の形態の第 1 の帳票例説明図である。

【図 3】本発明の一実施の形態の第 2 の帳票例説明図である。

【図 4】本発明の一実施の形態の全体フロー図（その

1）である。

【図 5】本発明の一実施の形態の全体フロー図（その 2）である。

【図 6】図 4 の罫線抽出処理の説明図である。

【図 7】図 4 のグループ抽出処理の説明図である。

【図 8】図 7 のグループ抽出処理の接続関係の説明図である。

【図 9】図 4 の表構成要素の分類処理の説明図である。

【図 10】図 4 の論理行分類処理説明図である。

【図 11】図 10 の論理行分類処理の動作説明図である。

【図 12】図 10 の入れ子抽出処理の説明図である。

【図 13】図 10 の表構造部の統合処理の説明図である。

【図 14】図 10 の罫線作成処理の説明図である。

【図 15】図 4 のブロック分類処理の説明図である。

【図 16】図 15 の罫線回復処理の説明図である。

【図 17】図 15 の論理行統合処理の説明図である。

【図 18】図 4 の論理列判定処理の説明図である。

【図 19】図 18 の列方向の入れ子抽出処理の説明図である。

【図 20】図 18 の 1 桁及び 3 桁罫線の入れ子構造の説明図である。

【図 21】図 18 の主表以外の箇所の入れ子解析の説明図である。

【図 22】図 4 のブロックの属性付与の説明図である。

【図 23】図 5 のフラグ設定処理の説明図である。

【図 24】図 5 の見出し認識処理の説明図である。

【図 25】図 5 の見出し文言辞書の説明図である。

【図 26】図 4 及び図 5 の処理のためのテーブルの関係図である。

【図 27】図 26 の横罫線テーブルの説明図である。

【図 28】図 26 の縦罫線テーブルの説明図である。

【図 29】図 26 の罫線項目テーブルの説明図である。

【図 30】斜度算出処理フロー図である。

【図 31】座標補正処理フロー図である。

【図 32】グループ処理フロー図である。

【図 33】図 32 のテーブル更新の説明図である。

【図 34】構成要素分類処理フロー図（その 1）である。

【図 35】構成要素分類処理フロー図（その 2）である。

【図 36】図 34 及び図 35 の分類後のテーブル説明図（その 1）である。

【図 37】図 34 及び図 35 の分類後のテーブル説明図（その 2）である。

【図 38】論理行分類処理フロー図である。

【図 39】入れ子構造解析処理フロー図である。

【図 40】論理行抽出後のテーブル説明図（その 1）である。

【図 4 1】論理行抽出後のテーブル説明図（その 2）である。

【図 4 2】統合処理フロー図である。

【図 4 3】帳票例 2 のテーブル説明図（その 1）である。

【図 4 4】帳票例 2 のテーブル説明図（その 2）である。

【図 4 5】帳票例 2 のグループ統合後のテーブル説明図（その 1）である。

【図 4 6】帳票例 2 のグループ統合後のテーブル説明図（その 2）である。

【図 4 7】罫線作成処理フロー図である。

【図 4 8】罫線作成の説明図である。

【図 4 9】罫線作成後のテーブル説明図（その 1）である。

【図 5 0】罫線作成後のテーブル説明図（その 2）である。

【図 5 1】ブロック分類処理フロー図である。

【図 5 2】ブロック分類後のテーブル説明図（その 1）である。

【図 5 3】ブロック分類後のテーブル説明図（その 2）である。

【図 5 4】罫線回復処理の説明図である。

【図 5 5】罫線回復処理フロー図である。

【図 5 6】罫線回復前のテーブル説明図（その 1）である。

【図 5 7】罫線回復前のテーブル説明図（その 2）である。

【図 5 8】罫線回復後のテーブル説明図（その 1）である。

【図 5 9】罫線回復後のテーブル説明図（その 2）である。

【図 6 0】罫線回復後のテーブル説明図（その 3）である。

【図 6 1】論理行統合処理の説明図である。

【図 6 2】論理行統合処理フロー図である。

【図 6 3】論理行統合前のテーブル説明図（その 1）である。

【図 6 4】論理行統合前のテーブル説明図（その 2）である。

【図 6 5】論理行統合後のテーブル説明図（その 1）である。

【図 6 6】論理行統合後のテーブル説明図（その 2）である。

【図 6 7】論理行統合後のテーブル説明図（その 3）である。

【図 6 8】論理列判定処理フロー図である。

【図 6 9】論理列判定後のテーブル説明図（その 1）である。

【図 7 0】論理列判定後のテーブル説明図（その 2）である。

【図 7 1】論理列判定後のテーブル説明図（その 3）である。

【図 7 2】1 桁、3 桁罫線の入れ子解析処理フロー図である。

【図 7 3】入れ子解析後のテーブル説明図である。

【図 7 4】主表以外の入れ子解析処理フロー図である。

【図 7 5】入れ子解析後のテーブル説明図（その 1）である。

【図 7 6】入れ子解析後のテーブル説明図（その 2）である。

【図 7 7】属性付与処理フロー図である。

【図 7 8】属性付与後のテーブル説明図である。

【図 7 9】フラグ設定処理フロー図である。

【図 8 0】フラグ設定後のテーブル説明図である。

【図 8 1】見出し認識処理フロー図である。

【図 8 2】見出し認識後のテーブル説明図である。

【図 8 3】属性設定処理フロー図である。

【図 8 4】属性設定後のテーブル説明図である。

【図 8 5】文字認識処理フロー図である。

【図 8 6】文字認識結果の説明図である。

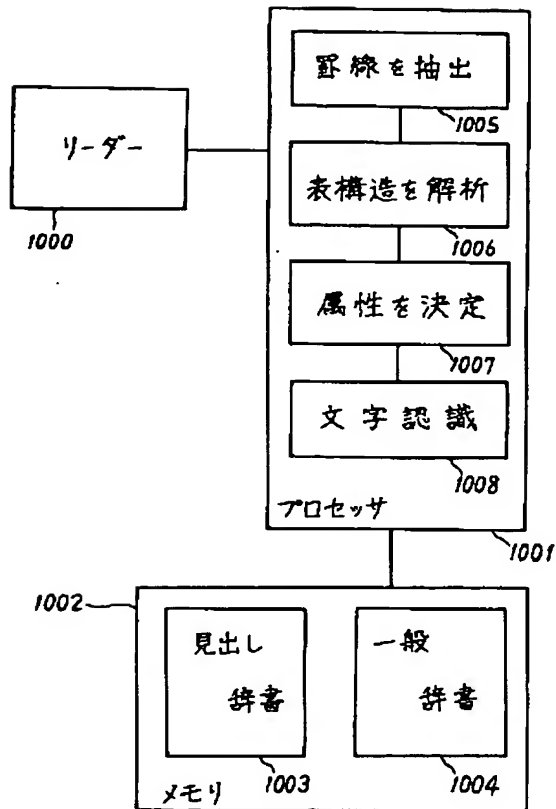
【図 8 7】従来技術の説明図である。

【符号の説明】

1000 リーダー
1001 プロセッサ
1002 メモリ
1003 見出し辞書
1004 一般辞書
1005 罫線抽出部
1006 表構造解析部
1007 属性決定部
1008 文字認識部

【図 1】

ブロック図



【図 2】

第1の帳票例説明図

振込依頼書

(1)	(2)	(3)				
(4)	振込人 (5)	2 3 4 5 6				
(7)						
(8)	(9)	(10)	(11)	(12)	(13)	(14)
(15)						
(16)	(17)	(18)	(19)	(20)	(21)	(22)
(23)						
(24)	(25)	(26)	(27)	(28)	(29)	(30)
(31)						
(32)	(33)	(34)	(35)	(36)	(37)	(38)
(39)						
(40)	(41)	(42)	(43)	(44)	(45)	(46)
(47)						
(48)	(49)	(50)	(51)	(52)	(53)	(54)
(55)						
(56)	(57)	(58)	(59)	(60)	(61)	(62)
(63)						
(64)	(65)	(66)	(67)	(68)	(69)	(70)
(71)						
(72)	(73)	(74)	(75)	(76)	(77)	(78)
(79)						
(80)	(81)	(82)	(83)	(84)	(85)	(86)
(87)						
(88)	(89)	(90)	(91)	(92)	(93)	(94)
(95)						
(96)	(97)	(98)	(99)	(100)	(101)	(102)
(103)						
(104)	(105)	(106)	(107)	(108)	(109)	(110)
(111)						
(112)	(113)	(114)	(115)	(116)	(117)	(118)
(119)						
(120)	(121)	(122)	(123)	(124)	(125)	(126)
(127)						
(128)	(129)	(130)	(131)	(132)	(133)	(134)
(135)						
(136)	(137)	(138)	(139)	(140)	(141)	(142)
(143)						
(144)	(145)	(146)	(147)	(148)	(149)	(150)
(151)						
(152)	(153)	(154)	(155)	(156)	(157)	(158)
(159)						
(160)	(161)	(162)	(163)	(164)	(165)	(166)
(167)						
(168)	(169)	(170)	(171)	(172)	(173)	(174)
(175)						
(176)	(177)	(178)	(179)	(180)	(181)	(182)
(183)						
(184)	(185)	(186)	(187)	(188)	(189)	(190)
(191)						
(192)	(193)	(194)	(195)	(196)	(197)	(198)
(199)						
(200)	(201)	(202)	(203)	(204)	(205)	(206)
(207)						
(208)	(209)	(210)	(211)	(212)	(213)	(214)
(215)						
(216)	(217)	(218)	(219)	(220)	(221)	(222)
(223)						
(224)	(225)	(226)	(227)	(228)	(229)	(230)
(231)						
(232)	(233)	(234)	(235)	(236)	(237)	(238)
(239)						
(240)	(241)	(242)	(243)	(244)	(245)	(246)
(247)						
(248)	(249)	(250)	(251)	(252)	(253)	(254)
(255)						
(256)	(257)	(258)	(259)	(260)	(261)	(262)
(263)						
(264)	(265)	(266)	(267)	(268)	(269)	(270)
(271)						
(272)	(273)	(274)	(275)	(276)	(277)	(278)
(279)						
(280)	(281)	(282)	(283)	(284)	(285)	(286)
(287)						
(288)	(289)	(290)	(291)	(292)	(293)	(294)
(295)						
(296)	(297)	(298)	(299)	(300)	(301)	(302)
(303)						
(304)	(305)	(306)	(307)	(308)	(309)	(310)
(311)						
(312)	(313)	(314)	(315)	(316)	(317)	(318)
(319)						
(320)	(321)	(322)	(323)	(324)	(325)	(326)
(327)						
(328)	(329)	(330)	(331)	(332)	(333)	(334)
(335)						
(336)	(337)	(338)	(339)	(340)	(341)	(342)
(343)						
(344)	(345)	(346)	(347)	(348)	(349)	(350)
(351)						
(352)	(353)	(354)	(355)	(356)	(357)	(358)
(359)						
(360)	(361)	(362)	(363)	(364)	(365)	(366)
(367)						
(368)	(369)	(370)	(371)	(372)	(373)	(374)
(375)						
(376)	(377)	(378)	(379)	(380)	(381)	(382)
(383)						
(384)	(385)	(386)	(387)	(388)	(389)	(390)
(391)						
(392)	(393)	(394)	(395)	(396)	(397)	(398)
(399)						
(400)	(401)	(402)	(403)	(404)	(405)	(406)
(407)						
(408)	(409)	(410)	(411)	(412)	(413)	(414)
(415)						
(416)	(417)	(418)	(419)	(420)	(421)	(422)
(423)						
(424)	(425)	(426)	(427)	(428)	(429)	(430)
(431)						
(432)	(433)	(434)	(435)	(436)	(437)	(438)
(439)						
(440)	(441)	(442)	(443)	(444)	(445)	(446)
(447)						
(448)	(449)	(450)	(451)	(452)	(453)	(454)
(455)						
(456)	(457)	(458)	(459)	(460)	(461)	(462)
(463)						
(464)	(465)	(466)	(467)	(468)	(469)	(470)
(471)						
(472)	(473)	(474)	(475)	(476)	(477)	(478)
(479)						
(480)	(481)	(482)	(483)	(484)	(485)	(486)
(487)						
(488)	(489)	(490)	(491)	(492)	(493)	(494)
(495)						
(496)	(497)	(498)	(499)	(500)	(501)	(502)
(503)						
(504)	(505)	(506)	(507)	(508)	(509)	(510)
(511)						
(512)	(513)	(514)	(515)	(516)	(517)	(518)
(519)						
(520)	(521)	(522)	(523)	(524)	(525)	(526)
(527)						
(528)	(529)	(530)	(531)	(532)	(533)	(534)
(535)						
(536)	(537)	(538)	(539)	(540)	(541)	(542)
(543)						
(544)	(545)	(546)	(547)	(548)	(549)	(550)
(551)						
(552)	(553)	(554)	(555)	(556)	(557)	(558)
(559)						
(560)	(561)	(562)	(563)	(564)	(565)	(566)
(567)						
(568)	(569)	(570)	(571)	(572)	(573)	(574)
(575)						
(576)	(577)	(578)	(579)	(580)	(581)	(582)
(583)						
(584)	(585)	(586)	(587)	(588)	(589)	(590)
(591)						
(592)	(593)	(594)	(595)	(596)	(597)	(598)
(599)						
(600)	(601)	(602)	(603)	(604)	(605)	(606)
(607)						
(608)	(609)	(610)	(611)	(612)	(613)	(614)
(615)						
(616)	(617)	(618)	(619)	(620)	(621)	(622)
(623)						
(624)	(625)	(626)	(627)	(628)	(629)	(630)
(631)						
(632)	(633)	(634)	(635)	(636)	(637)	(638)
(639)						
(640)	(641)	(642)	(643)	(644)	(645)	(646)
(647)						
(648)	(649)	(650)	(651)	(652)	(653)	(654)
(655)						
(656)	(657)	(658)	(659)	(660)	(661)	(662)
(663)						
(664)	(665)	(666)	(667)	(668)	(669)	(670)
(671)						
(672)	(673)	(674)	(675)	(676)	(677)	(678)
(679)						
(680)	(681)	(682)	(683)	(684)	(685)	(686)
(687)						
(688)	(689)	(690)	(691)	(692)	(693)	(694)
(695)						
(696)	(697)	(698)	(699)	(700)	(701)	(702)
(703)						
(704)	(705)	(706)	(707)	(708)	(709)	(710)
(711)						
(712)	(713)	(714)	(715)	(716)	(717)	(718)
(719)						
(720)	(721)	(722)	(723)	(724)	(725)	(726)
(727)						
(728)	(729)	(730)	(731)	(732)	(733)	(734)
(735)						
(736)	(737)	(738)	(739)	(740)	(741)	(742)
(743)						
(744)	(745)	(746)	(747)	(748)	(749)	(750)
(751)						
(752)	(753)	(754)	(755)	(756)	(757)	(758)
(759)						
(760)	(761)	(762)	(763)	(764)	(765)	(766)
(767)						
(768)	(769)	(770)	(771)	(772)	(773)	(774)
(775)						
(776)	(777)	(778)	(779)	(780)	(781)	(782)
(783)						
(784)	(785)	(786)	(787)	(788)	(789)	(790)
(791)						
(792)	(793)	(794)	(795)	(796)	(797)	(798)
(799)						
(800)	(801)	(802)	(803)	(804)	(805)	(806)
(807)						
(808)	(809)	(810)	(811)	(812)	(813)	(814)
(815)						
(816)	(817)	(818)	(819)	(820)	(821)	(822)
(823)						
(824)	(825)	(826)	(827)	(828)	(829)	(830)
(831)						
(832)	(833)	(834)	(835)	(836)	(837)	(838)
(839)						
(840)	(841)	(842)	(843)	(844)	(845)	(846)
(847)						
(848)	(849)	(850)	(851)	(852)	(853)	(854)
(855)						
(856)	(857)	(858)	(859)	(860)	(861)	(862)
(863)						
(864)	(865)	(866)	(867)	(868)	(869)	(870)
(871)						
(872)	(873)	(874)	(875)	(876)	(877)	(878)
(879)						
(880)	(881)	(882)	(883)	(884)	(885)	(886)
(887)						
(888)	(889)	(890)	(891)	(892)	(893)	(894)
(895)						
(896)	(897)	(898)	(899)	(900)	(901)	(902)
(903)						
(904)	(905)	(906)	(907)	(908)	(909)	(910)
(911)						
(912)	(913)	(914)	(915)	(916)	(917)	(918)
(919)						
(920)	(921)	(922)	(923)	(924)	(925)	(926)
(927)						
(928)	(929)	(930)	(931)	(932)	(933)	(934)
(935)						
(936)	(937)	(938)	(939)	(940)	(941)	(942)
(943)						
(944)	(945)	(946)	(947)	(948)	(949)	(950)
(951)						
(952)	(953)	(954)	(955)	(956)	(957)	(958)
(959)						
(960)	(961)	(962)	(963)	(964)	(965)	(966)
(967)						
(968)	(969)	(970)	(971)	(972)	(973)	(974)
(975)						
(976)	(977)	(978)	(979)	(980)	(981)	(982)
(983)						
(984)	(985)	(986)	(987)	(988)	(989)	(990)
(991)						
(992)	(993)	(994)	(995)	(996)	(997)	(998)
(999)						
(1000)	(1001)	(1002)	(1003)	(1004)	(1005)	(1006)
(1007)						
(1008)	(1009)	(1010)	(1011)	(1012)	(1013)	(1014)
(1015)						
(1016)	(1017)	(1018)	(1019)	(1020)	(1021)	(1022)
(1023)						
(1024)	(1025)	(1026)	(1027)	(1028)	(1029)	(1030)
(1031)						
(1032)	(1033)	(1034)	(1035)	(1036)	(1037)	(1038)
(1039)						
(1040)	(1041)	(1042)	(1043)	(1044)	(1045)	(1046)
(1047)						
(1048)	(1049)	(1050)	(1051)	(1052)	(1053)	(1054)
(1055)						
(1056)	(1057)	(1058)	(1059)	(1060)	(1061)	(1062)
(1063)						
(1064)	(1065)	(1066)	(1067)	(1068)	(1069)	(1070)
(1071)						
(1072)	(1073)	(1074)	(1075)	(1076)	(1077)	(1078)
(1079)						
(1080)	(1081)	(1082)	(1083)	(1084)	(1085)	(1086)
(1087)						
(1088)	(1089)	(1090)	(1091)	(1092)	(1093)	(1094)
(1095)						
(1096)	(1097)	(1098)	(1099)	(1100)	(1101)	(1102)
(1103)						
(1104)	(1105)	(1106)	(1107)	(1108)	(1109)	(1110)
(1111)						
(1112)	(1113)	(1114)	(1115)	(1116)	(1117)	(1118)
(1119)						
(1120)	(1121)	(1122)	(1123)	(1124)	(1125)	(1126)
(1127)						
(1128)	(1129)	(1130)	(1131)	(1132)	(1133)	(1134)
(1135)						
(1136)	(1137)	(1138)	(1139)	(1140)	(1141)	(1142)
(1143)						
(1144)	(1145)	(1146)	(1147)	(1148)	(1149)	(1150)
(1151)						
(1152)	(1153)	(1154)	(1155)	(1156)	(1157)	(1158)
(1159)						
(1160)	(1161)	(1162)	(1163)	(1164)	(1165)	(1166)
(1167)						
(1168)	(1169)	(1170)	(1171)	(1172)	(1173)	(1174)
(1175)						
(1176)	(1177)	(1178)	(1179)	(1180)	(1181)	(1182)
(1183)						
(1184)	(1185)	(1186)	(1187)	(1188)	(1189)	(1190)
(1191)						
(1192)	(1193)	(1194)	(1195)	(1196)	(1197)	(1198)
(1199)						
(1200)	(1201)	(1202)	(1203)	(1204)	(1205)	(1206)
(1207)						
(1208)	(1209)	(1210)	(1211)	(1212)	(1213)	(1214)
(1215)						
(1216)	(1217)	(1218)	(1219)	(1220)	(1221)	(1222)
(1223)						
(1224)	(1225)	(1226)	(1227)	(1228)	(1229)	(1230)
(1231)						
(1232)	(1233)	(1234)	(1235)	(1236)	(1237	

【図 4】

全体フロー図(その1)

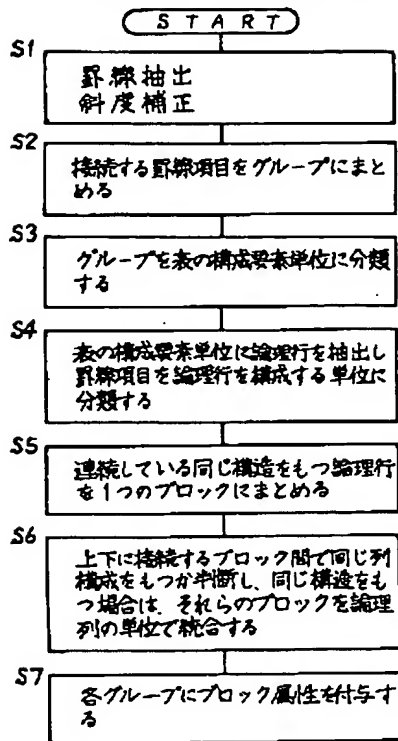
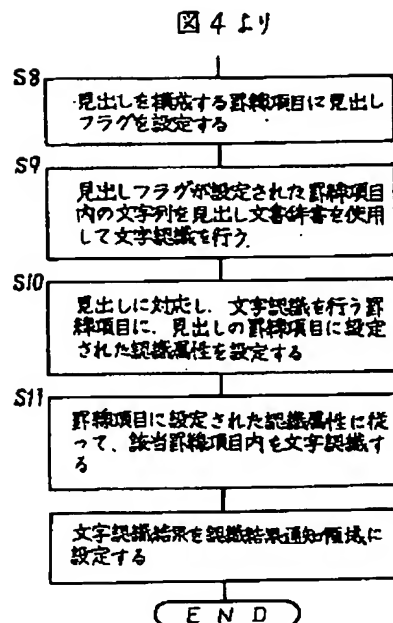


図 5へ

【図 5】

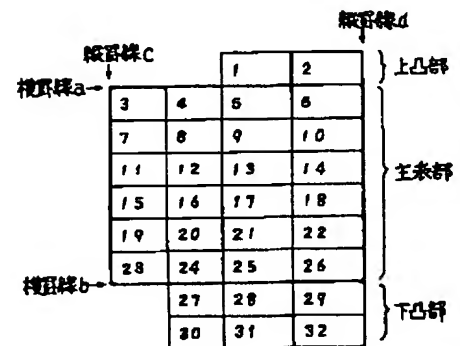
全体フロー図(その2)



【図 6】

【図 9】

表構成要素の分類処理の説明図



【図 12】

入れ子抽出処理の説明図

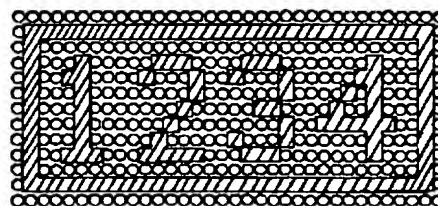
(A)

振込先		科目	口座番号
銀行名	支店名		
あいう銀行	かきく支店	普	12345678

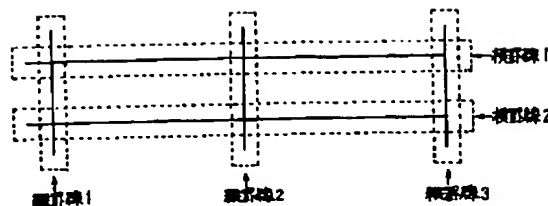
(B)

振込先 (罫線項目1)	
銀行名 (罫線項目2)	支店名 (罫線項目3)

(A)



(B)

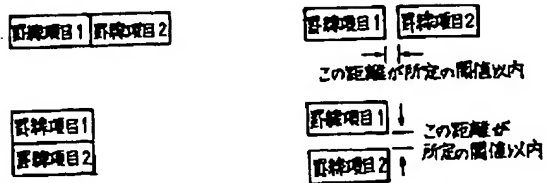


(C)

罫線項目1	罫線項目2
罫線項目3	罫線項目4

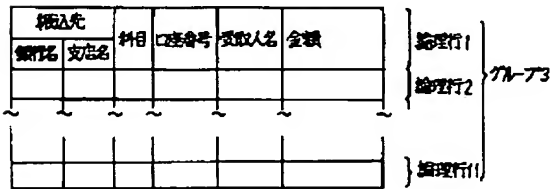
【図 8】

接続関係の説明図



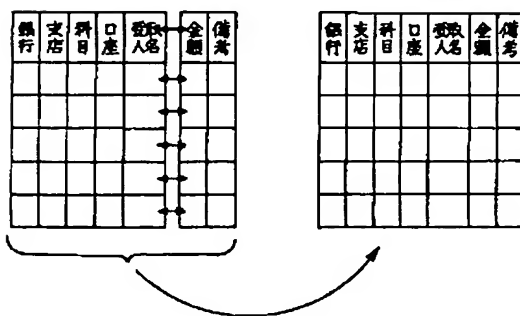
【図 11】

論理行分類処理の動作説明図



【図 13】

表構造部の統合処理の説明図

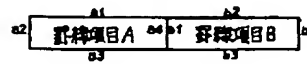


【図 22】

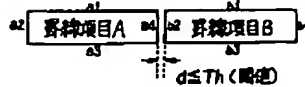
【図 10】

論理行分類処理説明図

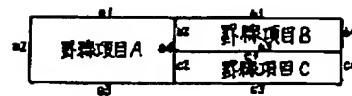
(A)



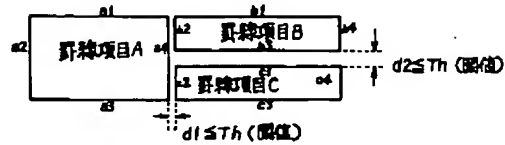
(B)



(C)

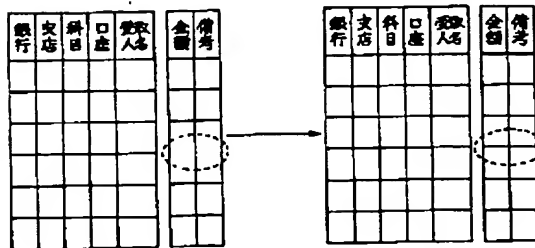


(D)



【図 14】

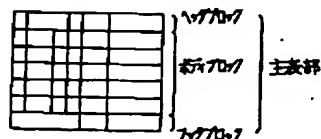
野線作成処理の説明図



【図 27】

横野線テーブルの説明図

ブロックの属性付与の説明図

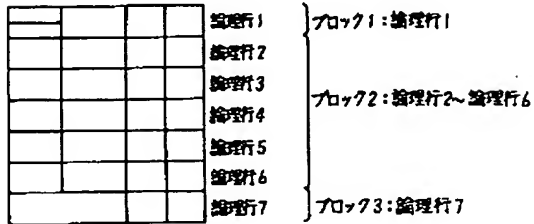


横野線テーブル

横野線番号	横野線属性	横野線値
1	Xh1, Yh1	Xh2, Yh2
2	Xh3, Yh3	Xh4, Yh4

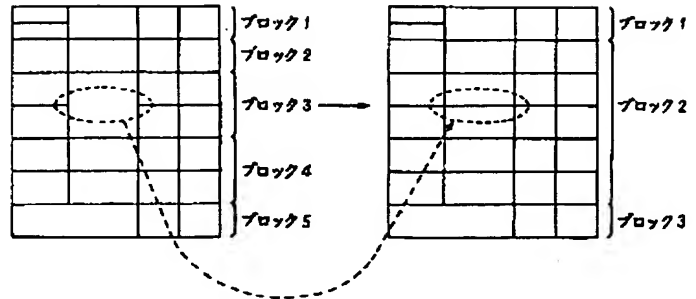
【図15】

ブロック分類処理の説明図



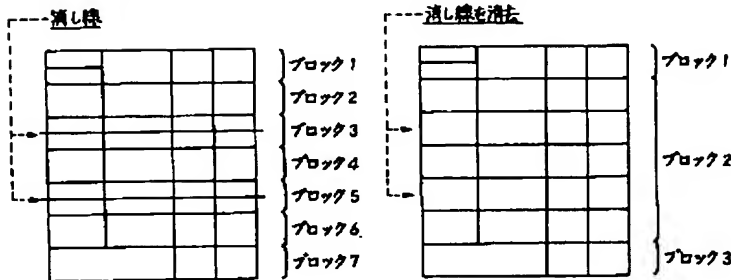
【図16】

罫線回復処理の説明図



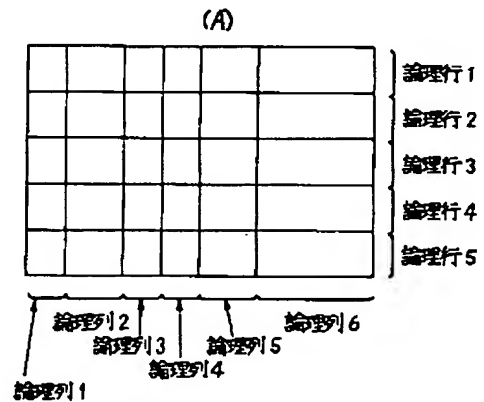
【図17】

論理行統合処理の説明図



【図19】

論理列判定処理の説明図



(B)

列方向の入れ子抽出処理の説明図

(A)

振込先 (振込先1)		科目
あいう銀行 (振込先2)	かきく支店 (振込先3)	普通

(B)

銀行名 (振込先1)	支店名 (振込先2)	科目
あいうえ銀行 かきくけ支店 (振込先3)		普通

振込先		科目
あいう銀行	かきく支店	普通

論理列1 論理列2

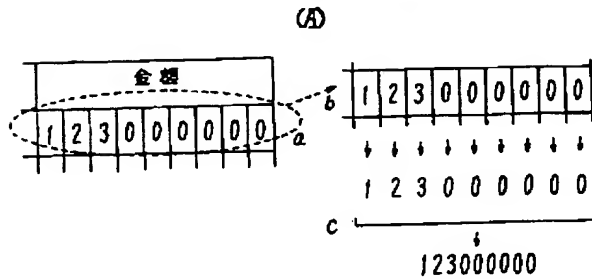
(C)

銀行名	支店名	科目
あいうえ銀行	かきくけ支店	普通

論理列1 論理列2

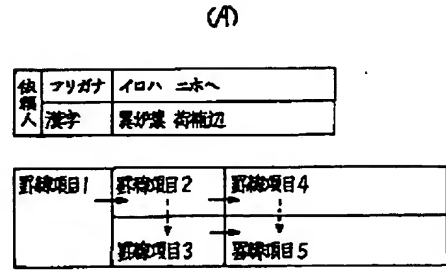
【図 20】

1桁及び3桁罫線の入れ子構造の説明図

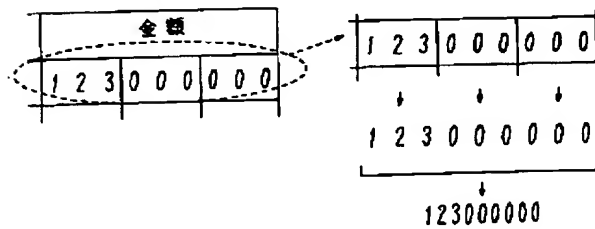


【図 21】

主表以外の箇所の入れ子解析の説明図

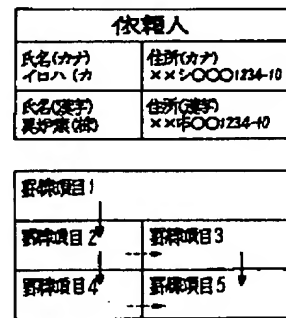


(B)

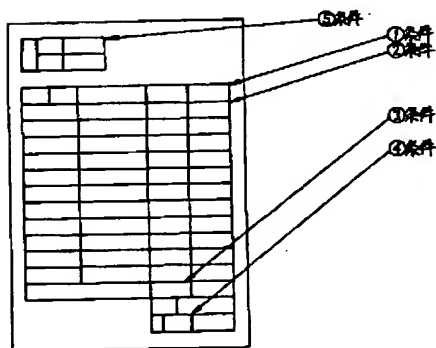


【図 23】

(B)



フラグ設定処理の説明図



【図 28】

【図 24】

見出し認識処理の説明図

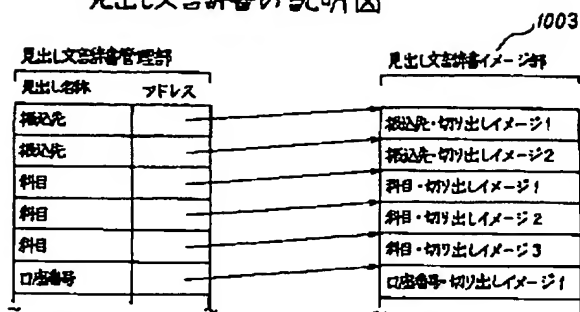
見出し名称	文字認識方式	文字フォント	他見出し リンク数	他見出しリンクアドレス
振込先	漢字	活字	2	銀行名のみ
銀行名	漢字	活字		
支店名	漢字	活字		
科目	漢字	活字		
口座番号	数字	活字		
金額	数字	手書き		

縦罫線テーブルの説明図

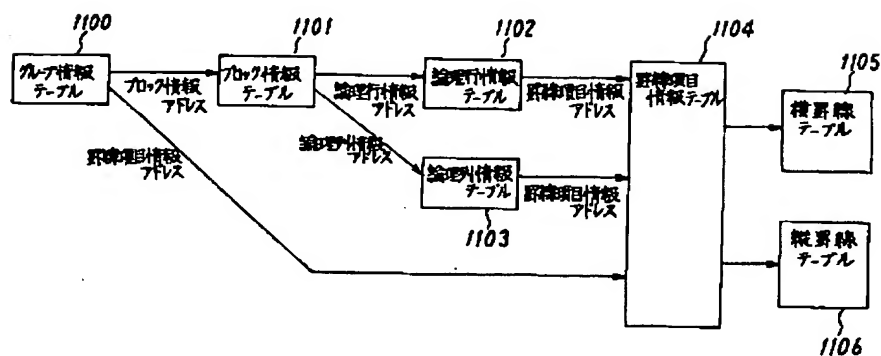
縦罫線テーブル

罫線項目	開始座標	終座標
1	Xv1, Yv1	Xv2, Yv2
2	Xv3, Yv3	Xv4, Yv4
3	Xv5, Yv5	Xv6, Yv6

見出し文言辞書の説明図



テーブルの関係図

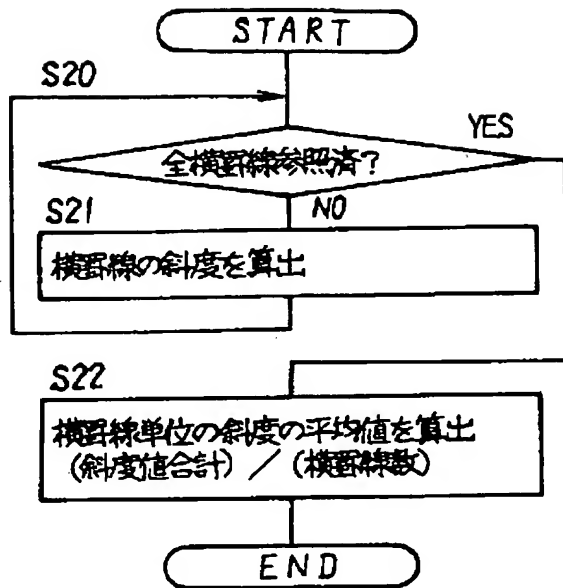


野線項目テーブルの説明図

界線項目情報テーブル															
座標番号 RM	座標座標 SC	座標座標 EC	グループ 番号 GN	フロー 番号 BN	行番号 LN	列番号 RO	上側 界線 番号 ULN	下側 界線 番号 LLN	左側 界線 番号 CPN	右側 界線 番号 RPN	上側 界線 項目 URN	下側 界線 項目 LRN	左側 界線 項目 ERN	右側 界線 項目 RRN	入水子 情報 IN
1	X1, Y1	X2, Y2					1	3	1	2					
2	X3, Y3	X4, Y4					1	2	2	3					

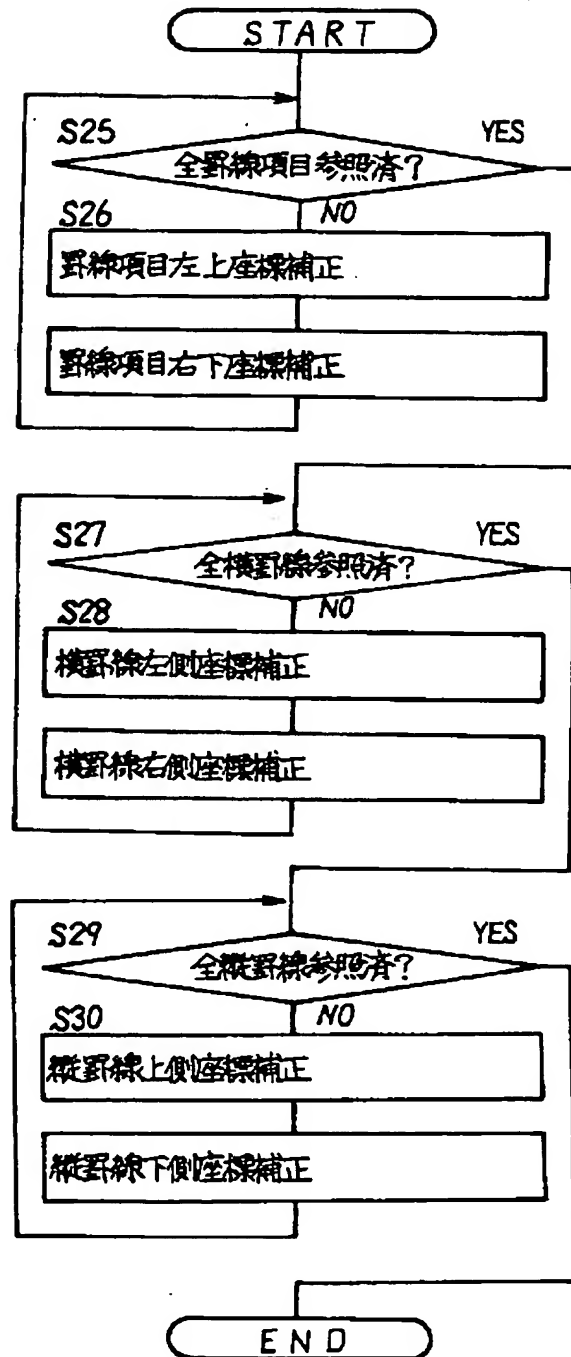
【図 30】

斜度算出処理フロー図



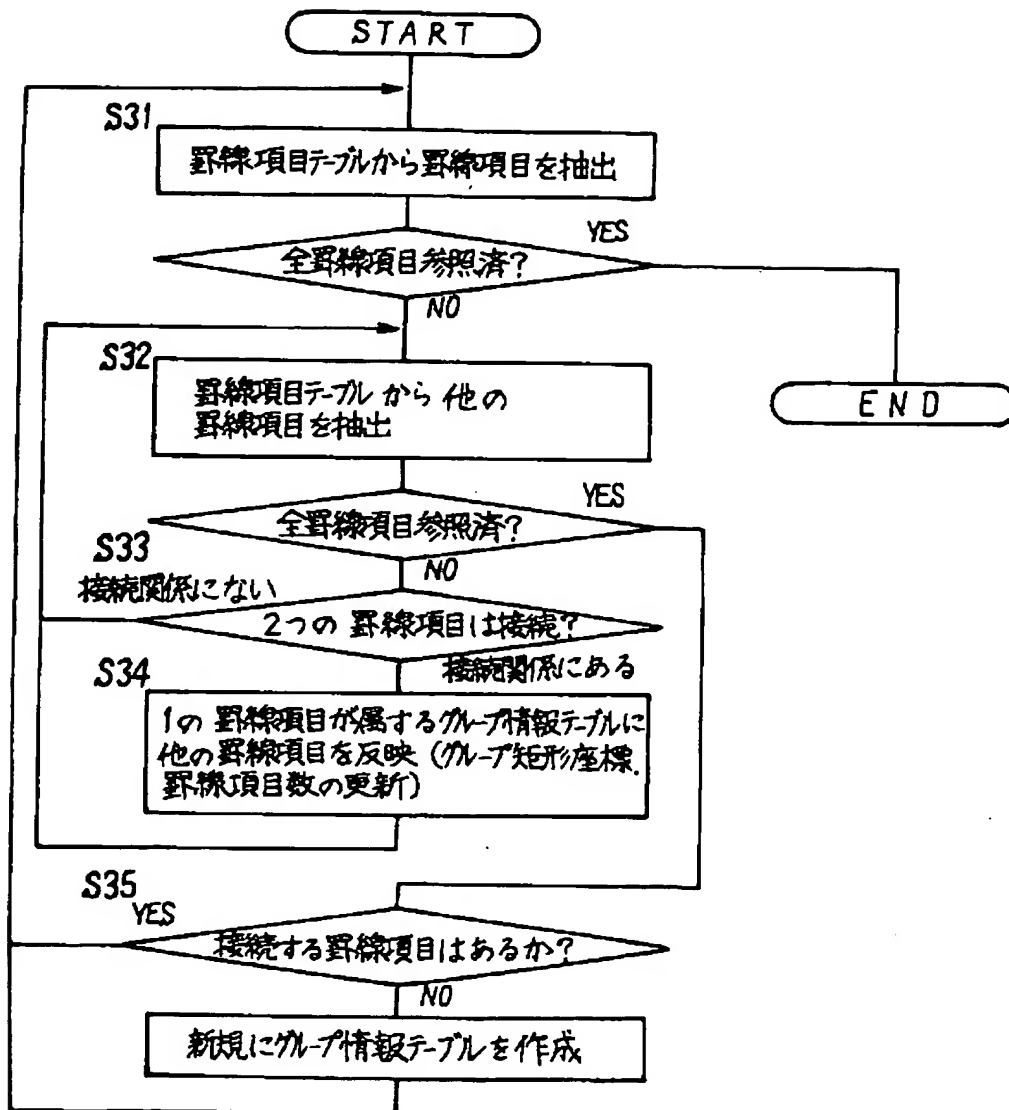
【図 31】

座標補正処理フロー図



【図 32】

グループ処理フロー図



【図 33】

テーブル更新の説明図

野線項目情報テーブル

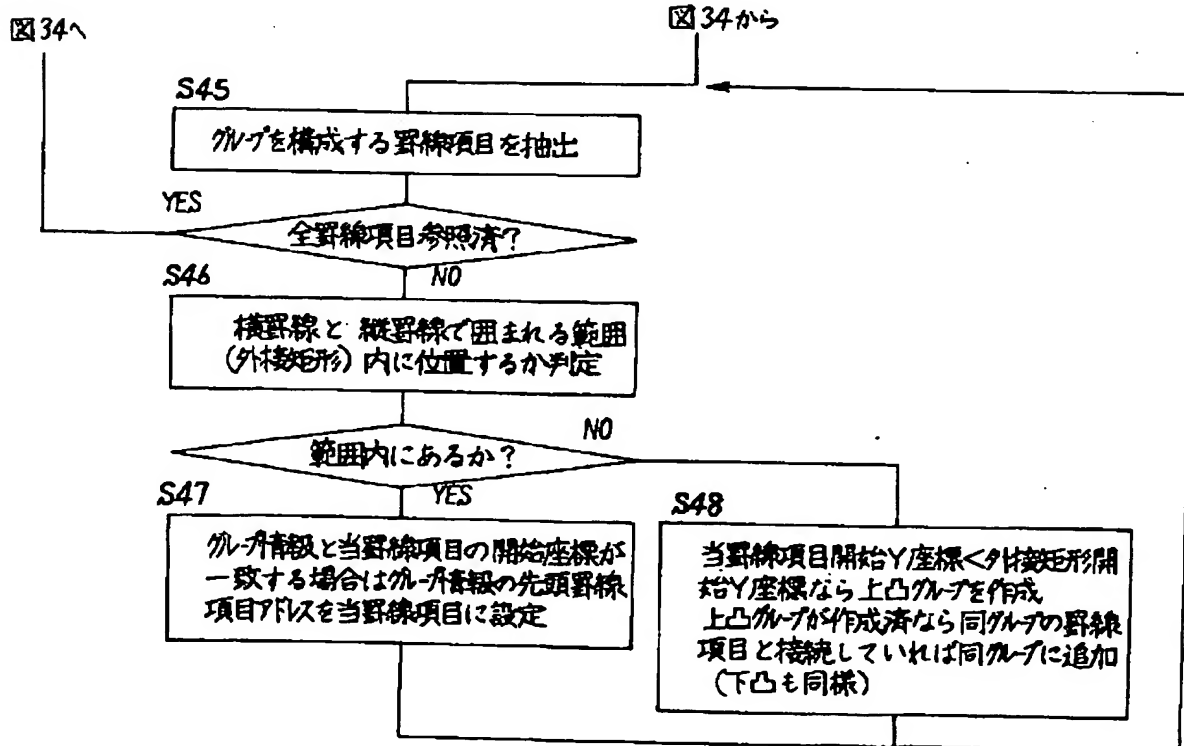
RN	EC	EC	GN	BN	LN	RO	ULN	LLN	LWN	RVN	URN	LRN	ERN	RRN	IN
1	X1,Y1	X2,Y2	1				1	3	1	2					
2	X3,Y3	X4,Y4	1				1	2	2	3					

グループ情報テーブル

グループ番号	開始座標	終了座標	グループ名	属性	先頭野線項目アドレス	先頭縦線項目アドレス	先頭横線項目アドレス
GN	EC	EC	RIN	BLN	LR	H1A	H2A
1	X1,Y1	X2,Y2	5			野線項目1アドレス	
2	X3,Y3	X4,Y4	171			野線項目6アドレス	

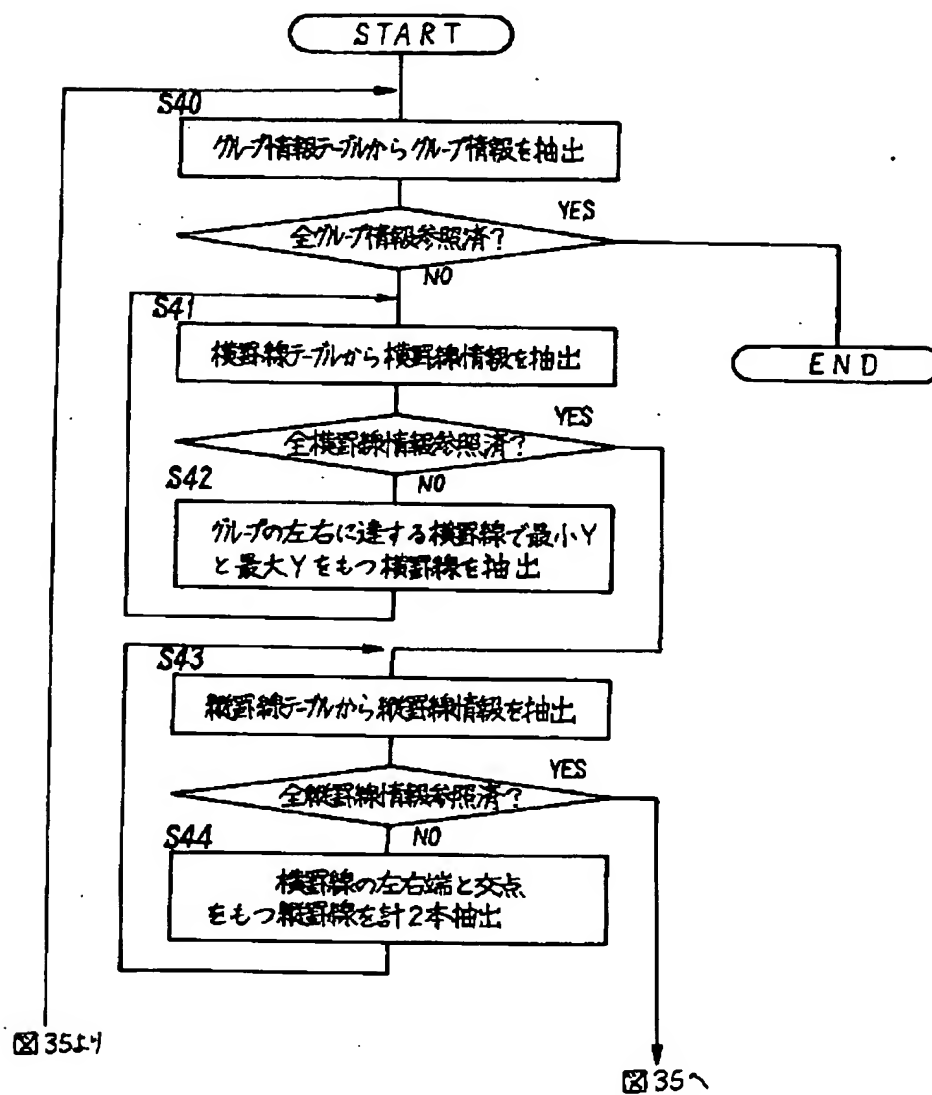
【図 35】

構成要素分類処理フロー図(その2)



【図34】

構成要素分類処理フロー図(その1)



【図 36】

分類後のテーブル説明図(その1)

詳細項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ULN	LLN	LVN	RVN	URN	LRN	ERN	RRN	IN
1	X1,Y1	X2,Y2	1				1	3	1	2					
2	X3,Y3	X4,Y4	1				1	2	2	3					
6	Xa,Ya	Xb,Yb	2				4	5	5	6					
10	Xc,Yc	Xd,Yd	3				6	7	8	10					
157	Xe,Ye	Xf,Yf	4				18	19	12	13					

【図 37】

分類後のテーブル説明図(その2)

グループ情報テーブル

GN	SC	EC	RIN	BLN	LR	HLA	HRA	HBA
1	X1,Y1	X2,Y2	5			野良項目17FVZ		
2	X3,Y3	X4,Y4	4			野良項目67FVZ		
3	X5,Y5	X6,Y6	147			野良項目107FVZ		
4	X7,Y7	X8,Y8	20			野良項目1577FVZ		

【図 43】

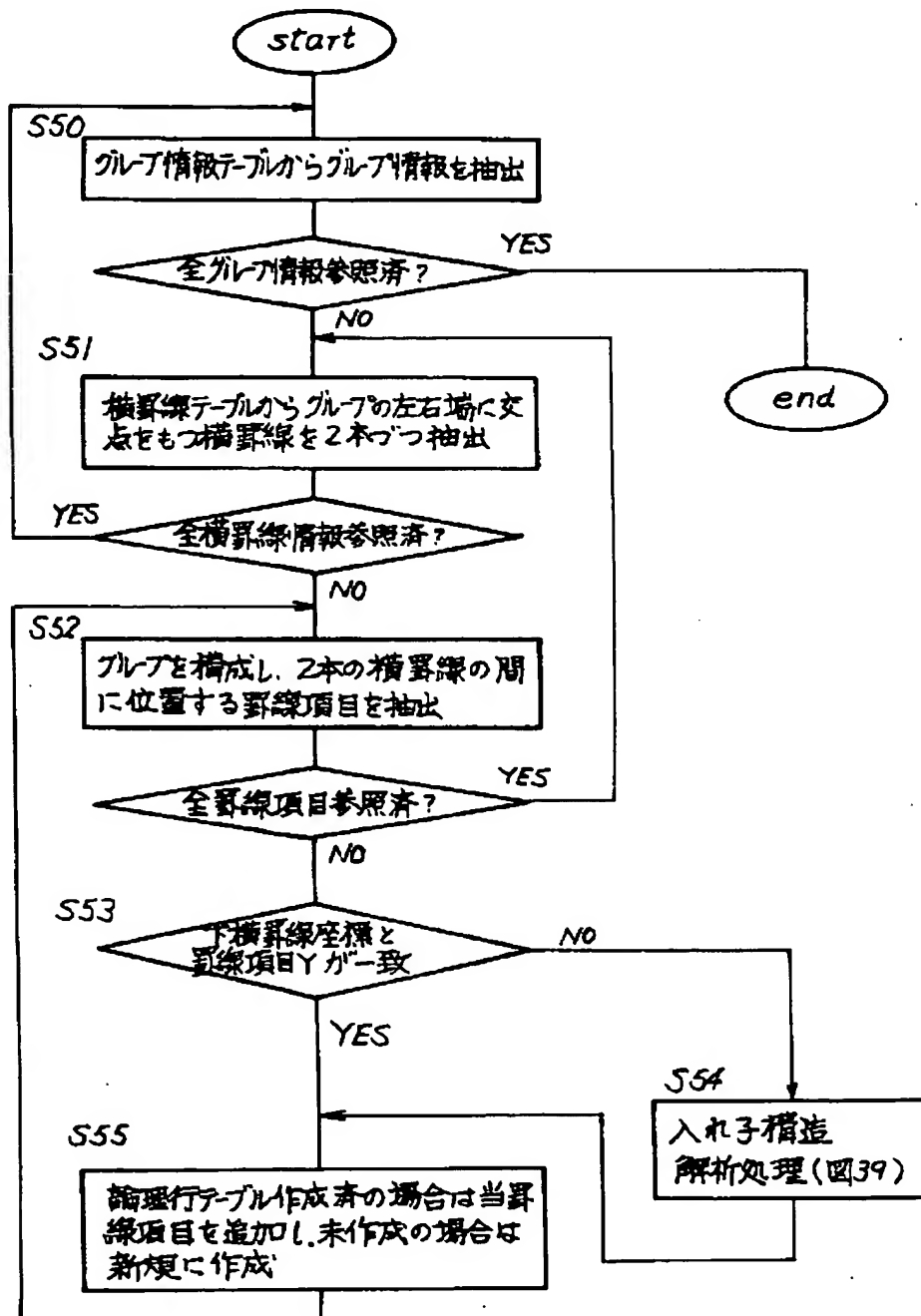
帳票例 2 の テーブル説明図 (その1)

詳細項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ULN	LLN	LVN	RVN	URN	LRN	ERN	RRN	IN
1	X1,Y1	X2,Y2	1		1		1	3	1	2					
6	X3,Y3	X4,Y4	2		1		4	5	5	6					ON
7	Xa,Ya	Xb,Yb	2		1		5	6	5	6					ON
132	Xc,Yc	Xd,Yd	3		1		16	17	8	9					
150	Xe,Ye	Xf,Yf	4		1		19	20	18	19					

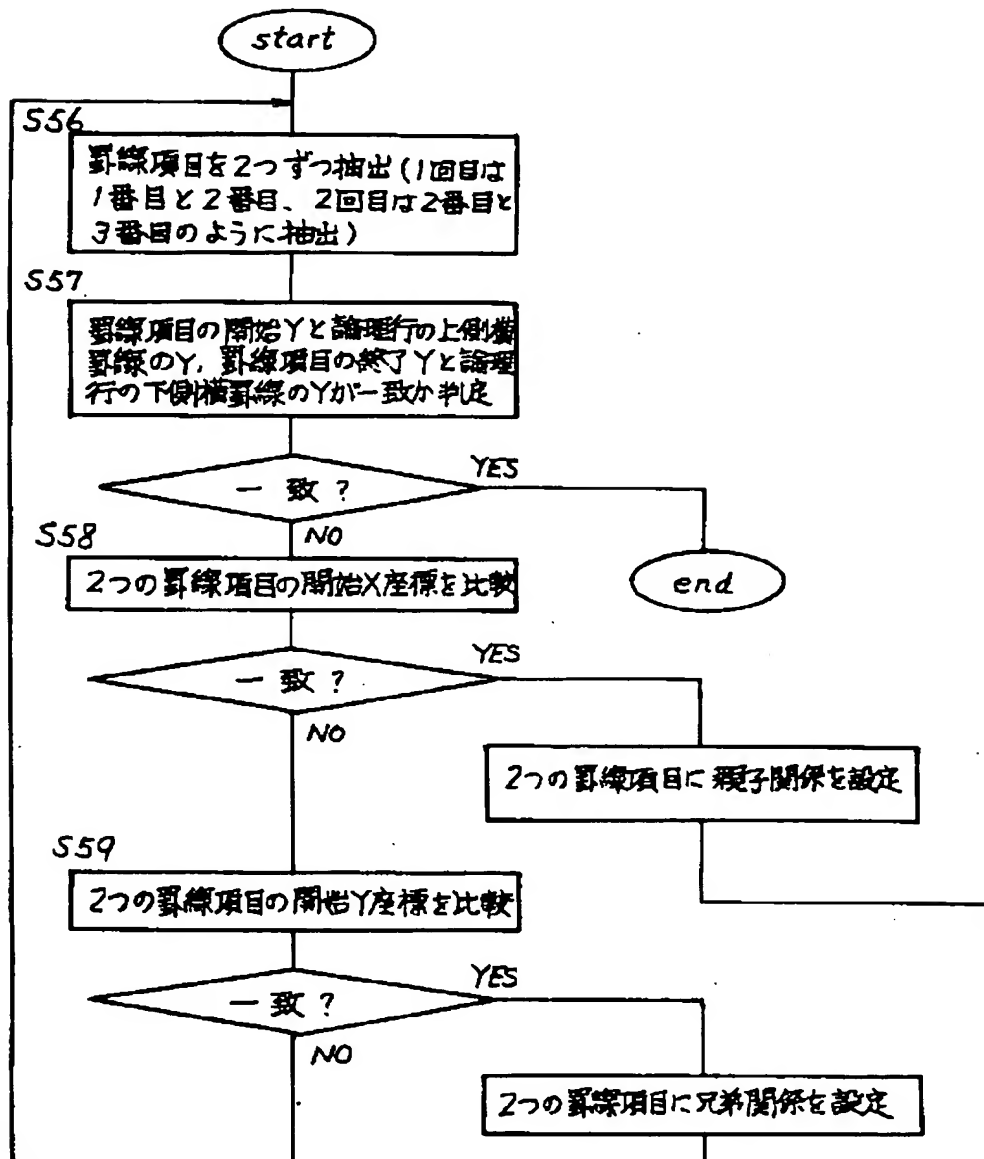
【図38】

論理行分類処理フロー図



【図39】

入れ子構造解析処理フロー図



【図 40】

論理行抽出後のテーブル説明図 (その1)

経緯項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ULN	LLN	LVN	RVN	URN	LRN	ERN	RRN	LN
1	X1.Y1	X2.Y2	1		1		1	3	1	2					
2	X3.Y3	X4.Y4	1		1		1	2	2	3					
6	Xa.Ya	Xb.Yb	2		1		4	5	5	6					
10	Xc.Yc	Xd.Yd	3		1		6	7	8	10		11			
11	Xe.Ye	Xf.Yf	3		1		6	7	8	10	10				12
12	Xg.Yg	Xh.Yh	3		1		6	7	8	10			11		
157	Xi.Yi	Xj.Yj	4		1		18	19	12	13					

RN	INF	KN	CRN	RLN
10	ON	0	1	0
11	ON	1		0
12	ON	1		1

【図 41】

論理行抽出後のテーブル説明図 (その2)

論理行情報テーブル

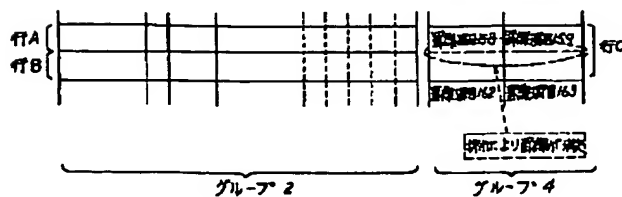
LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ELN	RRN	HLA
1	X1.Y1	X2.Y2	1		5	2	1	3	1	4	経緯項目1のアドレス
1	X3.Y3	X4.Y4	2		2	2	4	5	5	7	経緯項目6のアドレス
1	Xa.Ya	Xb.Yb	3		7	5	6	8	8	7	経緯項目10のアドレス
2	Xc.Yc	Xd.Yd	3		14	14	8	9	8	7	経緯項目17のアドレス
1	Xe.Ye	Xf.Yf	4		10	10	18	19	12	7	経緯項目157のアドレス

グループ情報テーブル

GN	SC	EC	RIN	BLN	LR	NLA	HRA	HBA
1	X1.Y1	X2.Y2	5		1	経緯項目17アドレス	経緯項目17アドレス	
2	X3.Y3	X4.Y4	4		2	経緯項目6アドレス	経緯項目27アドレス	
3	X5.Y5	X6.Y6	147		11	経緯項目107アドレス	経緯項目147アドレス	
4	X7.Y7	X8.Y8	20		2	経緯項目1577アドレス	経緯項目1577アドレス	

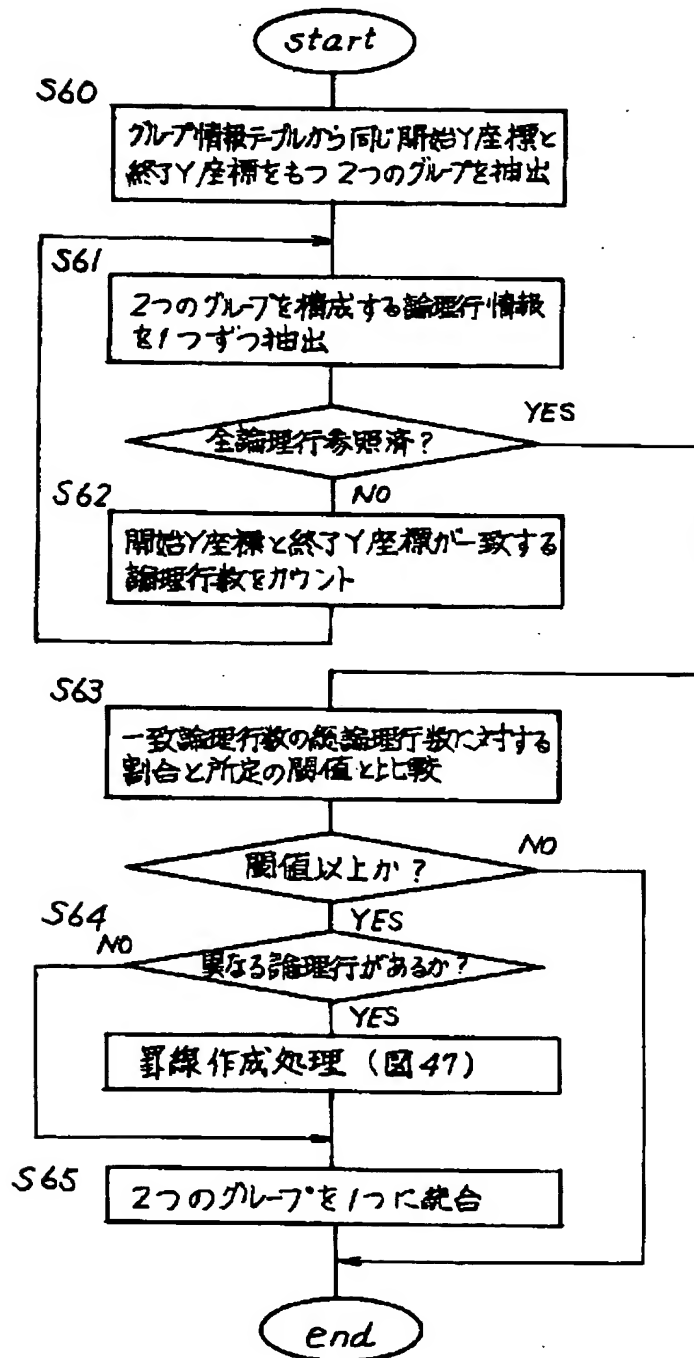
【図 48】

罫線作成の説明図



【図 42】

統合処理フロー図



【図 44】

帳票例2のテーブル説明図 (その2)

基準情報テーブル

LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ELN	RRN	HLA
1	X1.Y1	X2.Y2	1		5	3	1	3	1	4	基準項目1の7桁
1	X2.Y3	X4.Y4	2		6	5	4	6	5	17	基準項目6の7桁
2	X2.Y2	X6.Y6	2		12	12	6	7	5	17	基準項目12の7桁
1	Xc.Yc	Xd.Yd	3		9	9	16	17	8	17	基準項目32の7桁
1	Xe.Ye	Xf.Yf	4		2	2	19	20	18	20	基準項目157の7桁

グループ情報テーブル

GN	SC	EC	RIN	BLN	ER	HLA	HRA	HBA
1	X1.Y1	X2.Y2	5		1	基準項目177桁	基準項目77桁	
2	X3.Y3	X4.Y4	126		11	基準項目67桁	基準項目72桁	
3	X5.Y5	X6.Y6	18		2	基準項目132桁	基準項目37桁	
4	X7.Y7	X8.Y8	22		11	基準項目180桁	基準項目57桁	

【図 45】

帳票例2のグループ統合後のテーブル説明図 (その1)

基準項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ULN	LLN	LVN	RVN	URN	LRN	ERN	RRN	LN
1	X1.Y1	X2.Y2	1		1		1	3	1	2					
6	X3.Y3	X4.Y4	2		1		4	5	5	6					ON
7	X2.Y2	X6.Y6	2		1		5	6	5	6					ON
150	X2.Y2	X5.Y5	2		1		19	20	18	19					

指標情報テーブル

LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ELN	RRN	HLA
1	X1.Y1	X2.Y2	1		5	3	1	3	1	4	基準項目107桁
1	X3.Y3	X4.Y4	2		8	7	4	6	5	20	基準項目607桁
2	X2.Y2	X6.Y6	2		14	14	6	7	5	20	基準項目1807桁
1	Xe.Ye	Xf.Yf	4		2	2	19	20	18	20	基準項目1577桁

→ 前1段

【図46】

帳票例2のグループ統合後のテーブル説明図(その2)

グループ情報テーブル

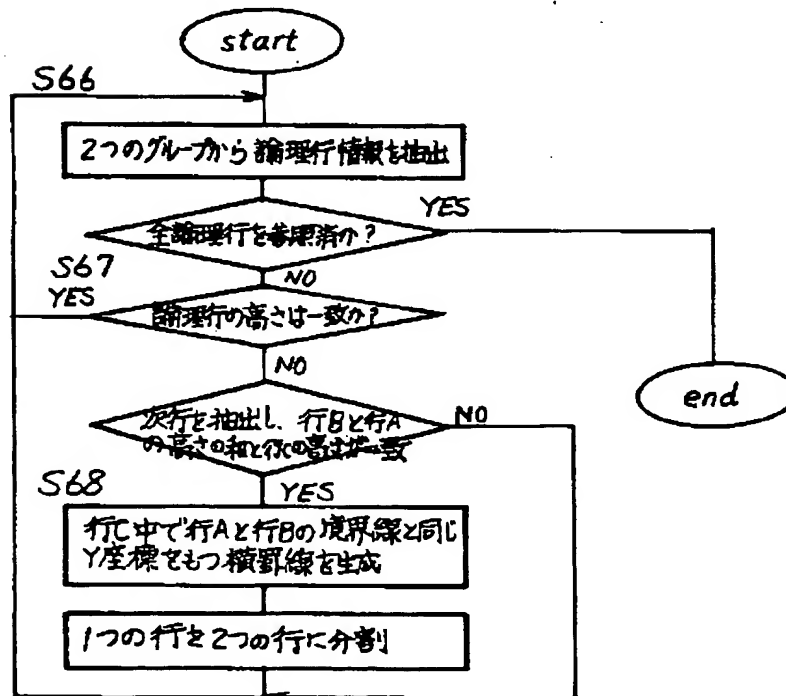
GN	SC	EC	RIN	BLN	LR	HLA	HRN	HBA
1	X1,Y1	X2,Y2	5		1	第1項目177F12	第2項目177F12	
2	X3,Y3	X4,Y4	48		11	第3項目677F12	第4項目277F12	
3	X5,Y5	X6,Y6	18		2	第5項目1327F12	第6項目777F12	
4	X7,Y7	X8,Y8	25		11	第7項目677F12	第8項目577F12	←削除

境界線情報テーブル

HN	SC	EC
4	X1,Y1	X2,Y2
5	X3,Y3	X4,Y4
6	X5,Y5	X6,Y6

【図47】

罫線作成処理フロー図



【図49】

罫線作成後のテーブル説明図 (その1)

罫線項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ULN	LLN	LVN	RVN	URN	LRN	ERN	RRN	IN
158	X1.Y1	X2.Y2	4		5		23	24	18	19					
159	X3.Y3	X4.Y2	4		5		23	24	19	20					
160	X1.Y1	X2.Y2	4		6		24	25	18	19					
161	X3.Y3	X4.Y2	4		6		24	25	19	20					
162	Xa.Ya	Xb.Yb	4		7		25	26	18	19					

【図50】

罫線作成後のテーブル説明図 (その2)

論理行情報テーブル

LN	SC	EC	GN	BN	BLN	LL	ULN	LLN	ELN	RRN	HLA
5	X1.Y1	X2.Y2	2		2	2	23	24	18	20	罫線項目5807Y2
6	X3.Y3	X4.Y4	2		2	2	23	24	18	20	罫線項目6007Y2

グループ情報テーブル

GN	SC	EC	RIN	BLN	LR	HLA	HRA	HBA
1	X1.Y1	X2.Y2	6		1	罫線項目17Y2	罫線項目17Y2	
2	X3.Y3	X4.Y4	126		11	罫線項目67Y2	罫線項目72Y2	
3	X5.Y5	X6.Y6	18		2	罫線項目1327Y2	罫線項目137Y2	
4	X7.Y7	X8.Y8	22		11	罫線項目507Y2	罫線項目57Y2	

【図52】

ブロック分類後のテーブル説明図 (その1)

罫線項目情報テーブル

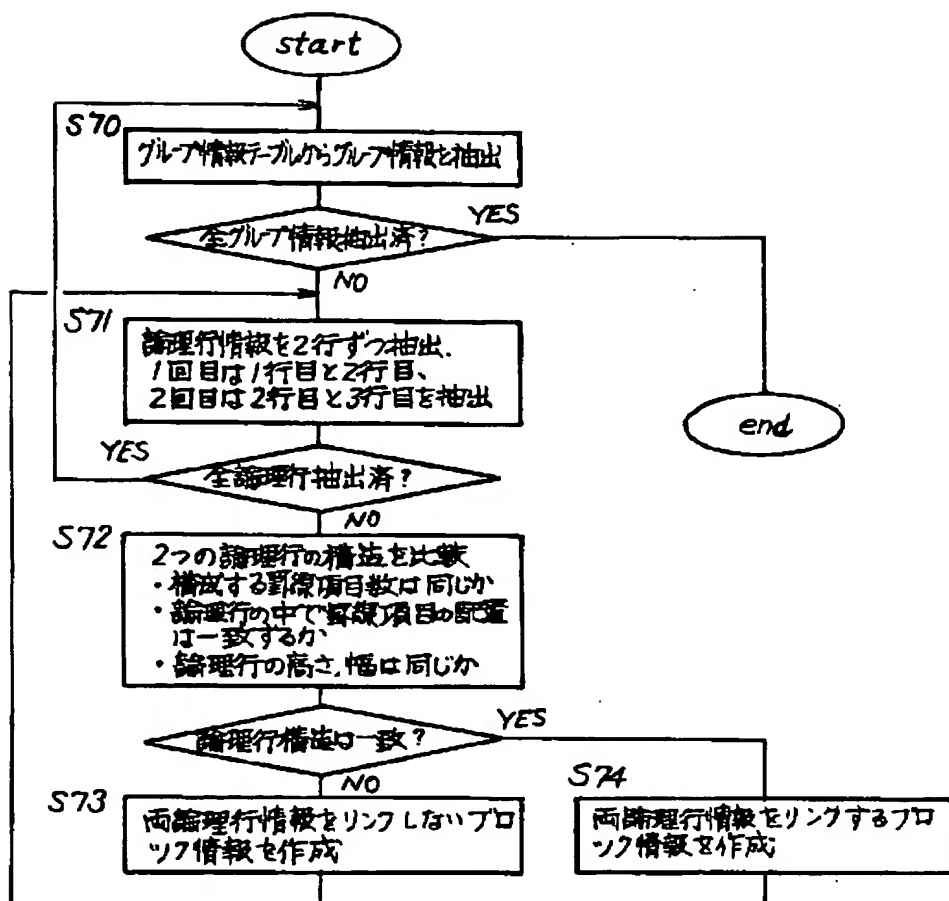
RN	SC	EC	GN	BN	LN	RO	ULN	LLV	LVN	RVN	URN	LRN	ERN	RRN	LN
1	X1.Y1	X2.Y2	1	1	1		1	3	1	2					
2	X3.Y3	X4.Y4	1	1	1		1	2	2	3					
6	Xa.Ya	Xb.Yb	2	1	1		4	5	5	6					
10	Xc.Yc	Xd.Yd	3	1	1		6	7	8	10				11	

論理行情報テーブル

LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ELN	RRN	HLA
1	X1.Y1	X2.Y2	1	1	5	2	1	3	1	4	罫線項目107Y2
1	Xa.Ya	Xb.Yb	3	1	7	5	6	8	8	7	罫線項目1007Y2
2	Xc.Yc	Xd.Yd	3	2	16	16	8	9	8	7	罫線項目1707Y2

【図 51】

ブロック分類処理フロー図



【図 54】

罫線回復処理の説明図

(10)	45	46	47	48	49	ブロックA
(11)	59	61	63	64	65	
(12)	60	62		(13)	66	ブロックC
(14)	85	86	87	88	89	
(15)						

【図 53】

ブロック分類後のテーブル説明図 (その2)

ブロック情報テーブル

BN	SC	EC	GN	RIN	LN	RO	HRA	HAA	ULN	LLN	LVN	RVN
1	X1.Y1	X2.Y2	1	5	1		論理項目17FVZ		1	3	1	4
1	X1.Y1	X2.Y2	2	4	2		論理項目27FVZ		4	6	5	7
1	X3.Y3	X4.Y4	3	7	1		論理項目47FVZ		6	8	8	7
2	X5.Y5	X6.Y6	3	140	10		論理項目57FVZ		8	18	8	7
1	X7.Y7	X8.Y8	4	20	2		論理項目57FVZ		18	20	12	7

グループ情報テーブル

GN	SC	EC	RIN	BLN	ER	HLA	HRA	HBA
1	X1.Y1	X2.Y2	5	1	1	論理項目17FVZ	論理項目17FVZ	論理項目17FVZ
2	X3.Y3	X4.Y4	4	1	2	論理項目67FVZ	論理項目27FVZ	論理項目27FVZ
3	X5.Y5	X6.Y6	147	2	11	論理項目107FVZ	論理項目47FVZ	論理項目37FVZ
4	X7.Y7	X8.Y8	20	1	2	論理項目57FVZ	論理項目57FVZ	論理項目57FVZ

【図 56】

【図 69】

罫線回復前のテーブル説明図 (その1)

罫線項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ULN	LLV	LVN	RVN	URN	LRN	ERN	RRN	IN
47	X1.Y1	X2.Y2	3	2	3		10	11	10	11					
48	X3.Y3	X4.Y4	3	2	3		10	11	11	12					
59	X5.Y5	X6.Y6	3	3	1		11	12	8	9					
60	X7.Y7	X8.Y8	3	3	1		12	14	8	9					
61	X9.Y9	X10.Y10	3	3	1		11	12	9	10					
62	X11.Y11	X12.Y12	3	3	1		12	14	9	10					
63	X13.Y13	X14.Y14	3	3	1		11	14	10	11					
64	X15.Y15	X16.Y16	3	3	1		11	14	11	12					

論理行情報テーブル

LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ERN	RRN	HLA
3	Xa.Ya	Xb.Yb	3	2	14	14	10	11	8	7	罫線項目8-45の7FVZ
1	Xc.Yc	Xd.Yd	3	3	26	14	11	14	8	7	罫線項目8-59の7FVZ
1	Xe.Ye	Xf.Yf	3	4	14	14	15	8	7	罫線項目8-63の7FVZ	

論理列判定後のテーブル説明図 (その1)

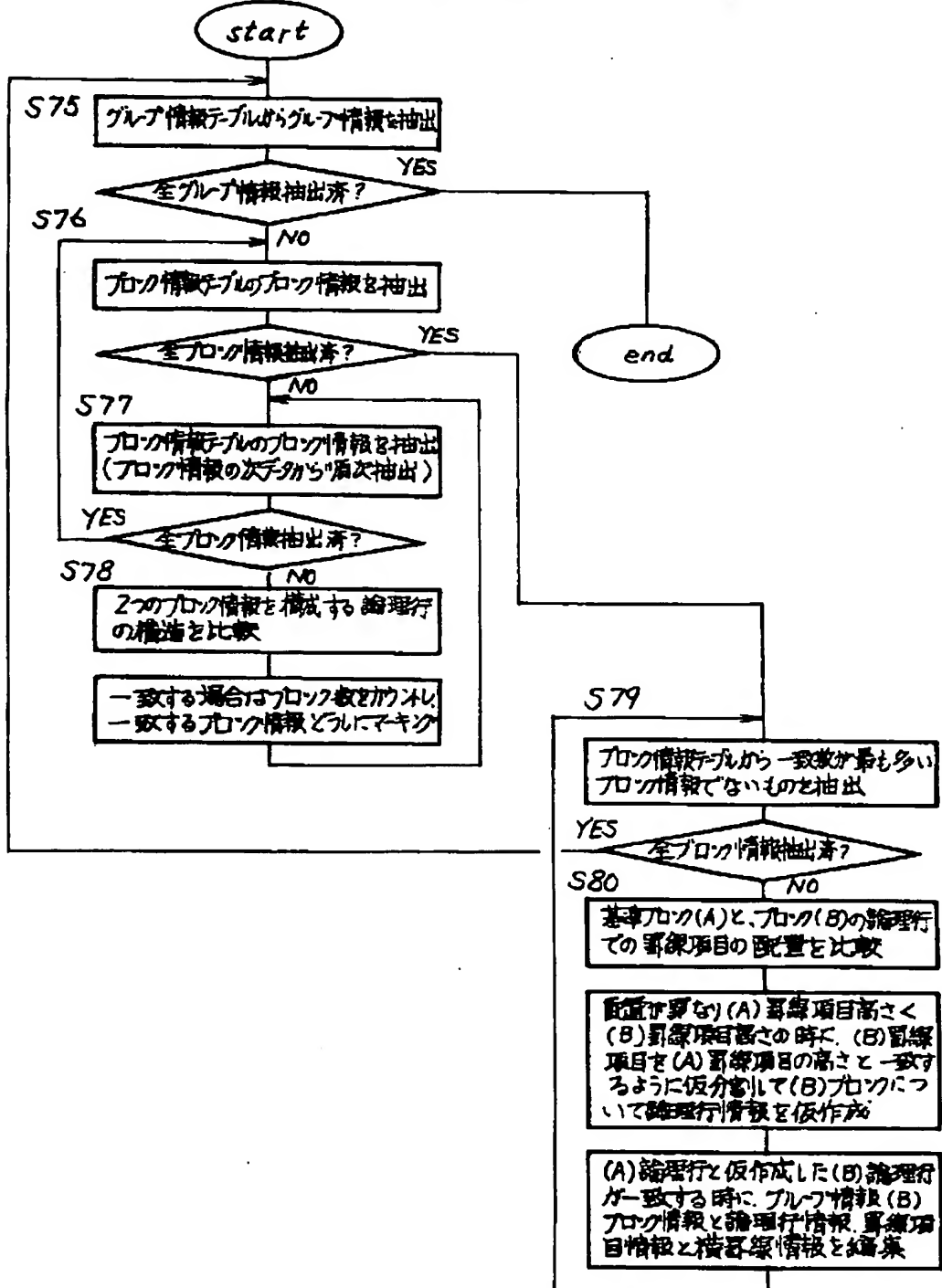
罫線項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ORN	LRN	ERN	RRN	IN	HAA
10	X1.Y1	X2.Y2	3	1	1	1		11				罫線項目8-17FVZ
11	X3.Y3	X4.Y4	3	1	1	1	10			12		罫線項目8-27FVZ
12	X5.Y5	X6.Y6	3	1	1	1			11			罫線項目8-37FVZ
17	X7.Y7	X8.Y8	3	2	1	1					ON	罫線項目8-47FVZ
18	X9.Y9	X10.Y10	3	2	1	1					ON	罫線項目8-57FVZ
19	X11.Y11	X12.Y12	3	2	1	2						罫線項目8-637FVZ
31	X13.Y13	X14.Y14	3	2	2	1					ON	罫線項目8-17FVZ
32	X15.Y15	X16.Y16	3	2	2	1					ON	罫線項目8-27FVZ
45	X17.Y17	X18.Y18	3	2	2	1					ON	罫線項目8-37FVZ
46	X19.Y19	X20.Y20	3	2	2	1					ON	罫線項目8-47FVZ

RN	INF	EN	CRN	RLN
17	ON	0		0
18	ON	0		1

【図55】

罫線回復処理フロー図



【図57】

罫線回復前のテーブル説明図(その2)

ブロック情報テーブル

BN	SC	EC	GN	RIN	LN	RO	HRA	HAA	LLN	LLN	LVN	RWN
1	X1.Y1	X2.Y2	3	7	1		論理行1TFIX		6	8	8	7
2	X3.Y3	X4.Y4	3	42	3		論理行2TFIX		8	11	8	7
3	X5.Y5	X6.Y6	3	26	1		論理行5TFIX		11	14	8	7
4	X7.Y7	X8.Y8	3	70	5		論理行6TFIX		15	19	8	7

グループ情報テーブル

GN	SC	EC	RIN	BLN	ER	HLA	HRA	HBA
3	X5.Y5	X6.Y6	145	4	10	罫線戻りNOTFIX	罫線戻りTFIX	ブロック3TFIX

【図58】

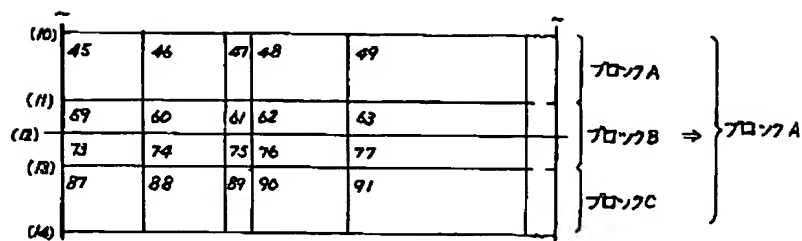
罫線回復後のテーブル説明図(その1)

罫線項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ULN	LLN	LVN	RWN	URN	LRN	ERN	RRN	EN
59	X5.Y5	X6.Y6	3	2	4		11	12	8	9					
61	Xa.Ya	Xb.Yb	3	2	4		11	12	9	10					
63	Xc.Yc	Xf.Yf	3	2	4		11	12	10	11					
64	Xg.Yg	Xh.Yh	3	2	4		11	12	11	12					
65	Xi.Yi	Xj.Yj	3	2	4		11	12	12	13					
60	X7.Y7	X8.Y8	3	2	5		12	14	8	9					
62	Xc.Yc	Xd.Yd	3	2	5		12	14	9	10					
177	Xk.Yk	Xl.Yl	3	2	5		12	14	10	11					
178	Xm.Ym	Xn.Yn	3	2	5		12	14	11	12					
66	Xo.Yo	Xp.Yp	3	2	5		12	14	12	13					
87	Xq.Yq	Xr.Yr	3	2	6		14	15	8	9					
88	Xs.Ys	Xt.Yt	3	2	6		14	15	9	10					

【図61】

論理行統合処理の説明図



【図 59】

罫線回復後のテーブル説明図 (その2)

論理行情報テーブル

LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ERN	RRN	HLA
4	X1.Y1	X2.Y2	3	2	14	14	11	12	8	7	罫線項目5907112
5	X3.Y3	X4.Y4	3	2	14	14	12	14	8	7	罫線項目6007112
6	X5.Y5	X6.Y6	3	2	14	14	14	15	8	7	罫線項目6507112
7	X7.Y7	X8.Y8	3	2	14	14	15	16	8	7	罫線項目9907112

ブロック情報テーブル

BN	SC	EC	GN	RIN	LN	RO	HRA	HAA	ULN	LLN	LVN	RVN
2	X1.Y1	X2.Y2	3	40	10		罫線項目27112		8	19	8	7
3	X3.Y3	X4.Y4	3	60	1		罫線項目37112		11	14	8	9
4	X5.Y5	X6.Y6	3	70	5		罫線項目57112		15	19	8	7

【図 60】

罫線回復後のテーブル説明図 (その3)

グループ情報テーブル

GN	SC	EC	RIN	BLN	ER	HLA	HRA	HBA
3	X3.Y3	X6.Y6	147	2	10	罫線項目307112	罫線項目47112	ブロック項目7112

抜き情報テーブル

HN	SC	EC
12	X1.Y1	X2.Y2
13	X3.Y3	X4.Y4

【図 71】

論理列判定後のテーブル説明図 (その3)

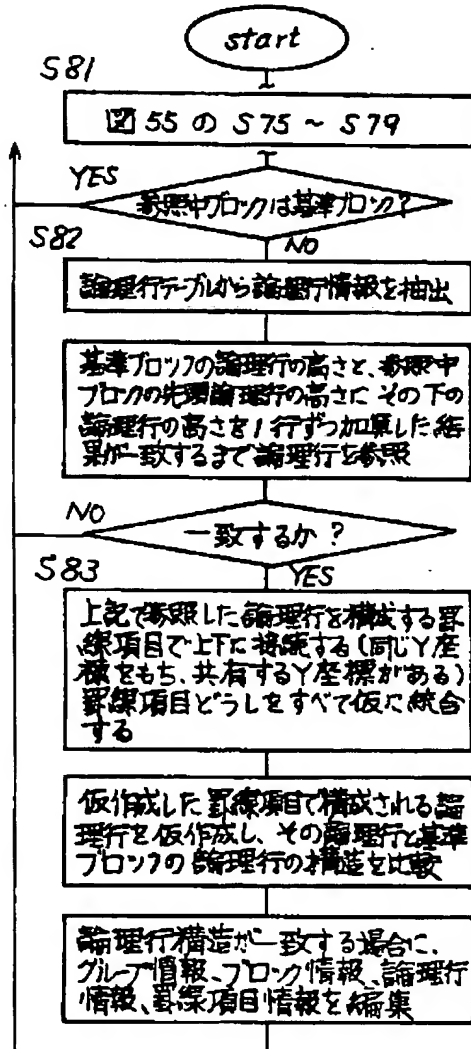
ブロック情報テーブル

BN	SC	EC	GN	RIN	LN	RO	HRA	HAA	ULN	LLN	LVN	RVN
1	X1.Y1	X2.Y2	1	5	1	3	罫線項目17112	罫線項目17112	1	3	1	4
1	X1.Y1	X2.Y2	2	4	2	2	罫線項目27112	罫線項目47112	4	6	5	7
1	X3.Y3	X4.Y4	3	7	1	5	罫線項目47112	罫線項目67112	6	8	8	7
2	X5.Y5	X6.Y6	3	140	10	5	罫線項目57112		8	18	8	7
1	X7.Y7	X8.Y8	4	20	2	10	罫線項目57112	罫線項目17112	18	20	12	7

【図 6 2】

【図 6 4】

論理行統合処理フロー図



論理行統合前のテーブル説明図 (その2)

論理行情報テーブル

LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ERN	RRN	HLA
3	Xa.Ya	Xb.Yb	3	2	14	14	10	11	8	7	基準項目507TFVZ
1	Xc.Yc	Xd.Yd	3	3	14	14	11	12	8	7	基準項目590TFVZ
2	Xa.Ya	Xf.Yf	3	3	14	14	12	13	8	7	基準項目730TFVZ

ブロック情報テーブル

BN	SC	EC	GN	RIN	LN	RO	HRA	HAA	ULN	LLN	LVN	RYN
1	X1.Y1	X2.Y2	3	7	1		基準項目7TFVZ		6	8	8	7
2	X3.Y3	X4.Y4	3	42	3		基準項目2TFVZ		8	11	8	7
3	X5.Y5	X6.Y6	3	28	2		基準項目5TFVZ		11	13	8	7
4	X7.Y7	X8.Y8	3	84	6		基準項目6TFVZ		13	19	8	7

グループ情報テーブル

GN	SC	EC	RIN	ULN	ER	HLA	HRA	HBA
3	X5.Y5	X6.Y6	161	4	12	基準項目810TFVZ	基準項目6TFVZ	ブロック37TFVZ

【図 7 0】

論理列判定後のテーブル説明図 (その2)

論理行情報テーブル

LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ELN	RRN	HLA
1	X1.Y1	X2.Y2	1	1	5	2	1	3	1	4	基準項目107TFVZ
1	Xa.Ya	Xb.Yb	3	1	7	5	6	8	8	7	基準項目1007TFVZ
2	Xc.Yc	Xd.Yd	3	2	14	5	8	9	8	7	基準項目1707TFVZ

論理列情報テーブル

RO	SC	EC	GN	BN	RIN	ULN	LLN	ELN	RRN	HLA
1	X1.Y1	X2.Y2	1	1	1	1	3	1	2	基準項目107TFVZ
1	X3.Y3	X4.Y4	3	1	23	6	18	8	10	基準項目1007TFVZ
2	X5.Y5	X6.Y6	3	1	11	6	18	10	11	基準項目1307TFVZ

【図 63】

論理行統合前のテーブル説明図(その1)

異種項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ULN	LLV	LNV	RNV	URN	LRN	ERN	RRN	IN
45	X1.Y1	X2.Y2	3	2	3		10	11	8	9					
46	X3.Y3	X4.Y4	3	2	3		10	11	9	10					
59	X5.Y5	X6.Y6	3	3	1		11	12	8	9					
60	X7.Y7	X8.Y8	3	3	1		11	12	9	10					
61	Xa.Ya	Xb.Yb	3	3	1		11	12	10	11					
73	Xc.Yc	Xd.Yd	3	3	2		12	13	8	9					
74	Xe.Ye	Xf.Yf	3	3	2		12	13	9	10					
75	Xg.Yg	Xh.Yh	3	3	2		12	13	10	11					

【図 65】

論理行統合後のテーブル説明図(その1)

異種項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	ULN	LLV	LNV	RNV	URN	LRN	ERN	RRN	IN
45	X1.Y1	X2.Y2	3	2	3		10	11	8	9					
46	X1.Y3	X4.Y4	3	2	3		10	11	9	10					
59	X5.Y5	X6.Y6	3	2	4		11	13	8	9					
60	X7.Y7	X8.Y8	3	2	4		11	13	9	10					
61	Xa.Ya	Xb.Yb	3	2	4		11	13	10	11					
73	Xc.Yc	Xd.Yd	3	3	2		12	13	8	9					
74	Xe.Ye	Xf.Yf	3	3	2		12	13	9	10					
75	Xg.Yg	Xh.Yh	3	3	2		12	13	10	11					

【図 66】

論理行統合後のテーブル説明図 (その2)

論理行テーブル

LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ERN	RBN	HLA
3	X1.Y1	X2.Y2	3	2	14	14	10	11	8	7	属性項目45のアドレス
4	X3.Y3	X4.Y4	3	2	14	14	11	13	8	7	属性項目30のアドレス
5	X5.Y5	X6.Y6	3	2	14	14	12	13	8	7	属性項目70のアドレス

ブロック情報テーブル

BN	SC	EC	GN	RIN	LN	RO	HRA	HAA	ULN	LLN	LVN	RVL
1	X1.Y1	X2.Y2	3	7	1		属性項目17のアドレス		6	8	8	7
2	X3.Y3	X4.Y4	3	10	10		属性項目27のアドレス		8	19	8	7
3	X5.Y5	X6.Y6	3	28	2		属性項目57のアドレス		11	13	8	7
4	X7.Y7	X8.Y8	3	36	6		属性項目67のアドレス		13	19	8	7

【図 67】

論理行統合後のテーブル説明図 (その3)

グループ情報テーブル

GN	SC	EC	RIN	BLN	ER	HLA	HRA	HBA
3	X5.Y5	X6.Y6	147	2	11	属性項目107のアドレス	属性項目47のアドレス	ブロック37のアドレス

横断線情報テーブル

HN	SC	EC
11	X1.Y1	X2.Y2
12	X3.Y3	X4.Y4

【図 82】

見出し認識後のテーブル説明図

属性項目情報テーブル

RN	GN	BN	IN	RO	INF	CC	CF	HAA
1	1	1	1	1	1	漢字	活字	属性項目27のアドレス
2	1	1	1	1	1	カナ	活字	属性項目37のアドレス
3	1	1	1	1	1			属性項目47のアドレス
4	1	1	1	1	1	漢字	活字	属性項目57のアドレス
5	1	1	1	1	1			
6	2	1	1	1	1	漢字	活字	属性項目87のアドレス
7	2	1	1	2	0			属性項目97のアドレス
8	2	1	2	1	1	漢字	活字	
9	2	1	2	2	0			
10	3	1	1	1	1	漢字	活字	属性項目117のアドレス
11	3	1	1	1	1	漢字	活字	属性項目117のアドレス
12	3	1	1	1	1	漢字	活字	属性項目117のアドレス
17	3	2	1	1	1	漢字	活字	属性項目187のアドレス
18	3	2	1	1	1			属性項目317のアドレス
157	4	1	1	1	1	数字	平書き	属性項目1877のアドレス
158	4	1	1	2	1			属性項目15977のアドレス

【図 84】

属性設定後のテーブル説明図

属性項目情報テーブル

RN	GN	BN	IN	RO	INF	CC	CF	HAA
1	1	1	1	1	1	漢字	活字	属性項目27のアドレス
2	1	1	1	1	1	カナ	活字	属性項目37のアドレス
3	1	1	1	1	1	カナ	活字	属性項目47のアドレス
4	1	1	1	1	1	漢字	活字	属性項目57のアドレス
5	1	1	1	1	1	漢字	活字	
6	2	1	1	1	1	漢字	活字	属性項目87のアドレス
7	2	1	1	2	0	漢字	活字	属性項目97のアドレス
8	2	1	2	1	1	漢字	活字	
9	2	1	2	2	0	漢字	活字	
10	3	1	1	1	1	漢字	活字	属性項目117のアドレス
11	3	1	1	1	1	漢字	活字	属性項目117のアドレス
12	3	1	1	1	1	漢字	活字	属性項目117のアドレス
17	3	2	1	1	1	漢字	活字	属性項目187のアドレス
18	3	2	1	1	1	漢字	活字	属性項目317のアドレス
157	4	1	1	1	1	数字	平書き	属性項目1877のアドレス
158	4	1	1	2	1	数字	平書き	属性項目15977のアドレス

【図68】

論理行統合後のテーブル説明図 (その2)

論理行テーブル

LN	SC	EC	GN	BN	RIN	LL	UIN	LLN	ERN	BRN	HLA
3	X4.Y4	X4.Y4	3	2	14	14	10	11	8	7	第4項目は4507112
4	X4.Y4	X4.Y4	3	2	14	14	11	13	8	7	第4項目は3907112
5	X4.Y4	X4.Y4	3	2	14	14	10	10	8	7	第4項目は7007112

ブロック情報テーブル

BN	SC	EC	GN	RIN	LN	RO	HRA	HAA	ULN	LLN	LVN	RVL
1	X4.Y1	X2.Y2	3	7	1		第1項目は717112		6	8	8	7
2	X3.Y3	X4.Y4	3	140	10		第1項目は727112		8	19	8	7
3	X4.Y4	X4.Y4	3	28	2		第1項目は57112		11	13	8	7
4	X7.Y7	X2.Y8	3	34	6		第1項目は57112		13	19	8	7

【図73】

【図86】

入れ子解析後のテーブル説明図

親項目情報テーブル

RN	SC	EC	GN	BN	LN	RO	IN					HAA
							INF	KN	CRM	RIN	AN	
16	X4.Y1	X2.Y2	3	2	1	5						第1項目は727112
22	X3.Y3	X4.Y4	3	2	1	5	ON	0	0	0	23	第1項目は237112
23	X5.Y5	X6.Y6	3	2	1	5	ON	0	0	1	24	第1項目は247112
24	X7.Y7	X8.Y8	3	2	1	5	ON	0	0	2	25	第1項目は257112
25	X4.Y4	X4.Y4	3	2	1	5	ON	0	0	3	26	第1項目は267112
26	X4.Y4	X4.Y4	3	2	1	5	ON	0	0	4	27	第1項目は277112
27	X4.Y4	X4.Y4	3	2	1	5	ON	0	0	5	28	第1項目は287112
28	X4.Y4	X4.Y4	3	2	1	5	ON	0	0	6	29	第1項目は297112
29	X4.Y4	X4.Y4	3	2	1	5	ON	0	0	7	30	第1項目は307112
30	X4.Y4	X4.Y4	3	2	1	5	ON	0	0	8	0	第1項目は317112
36	X4.Y4	X4.Y4	3	2	2	5	ON	0	0	0	37	第1項目は377112
37	X4.Y4	X4.Y4	3	2	2	5	ON	0	0	0	38	第1項目は387112

文字認識結果の説明図

認識結果 通知帳管理用 / 200

認識結果見出し数: 13
認識結果見出し形式アドレス

認識結果見出し領域 / 201

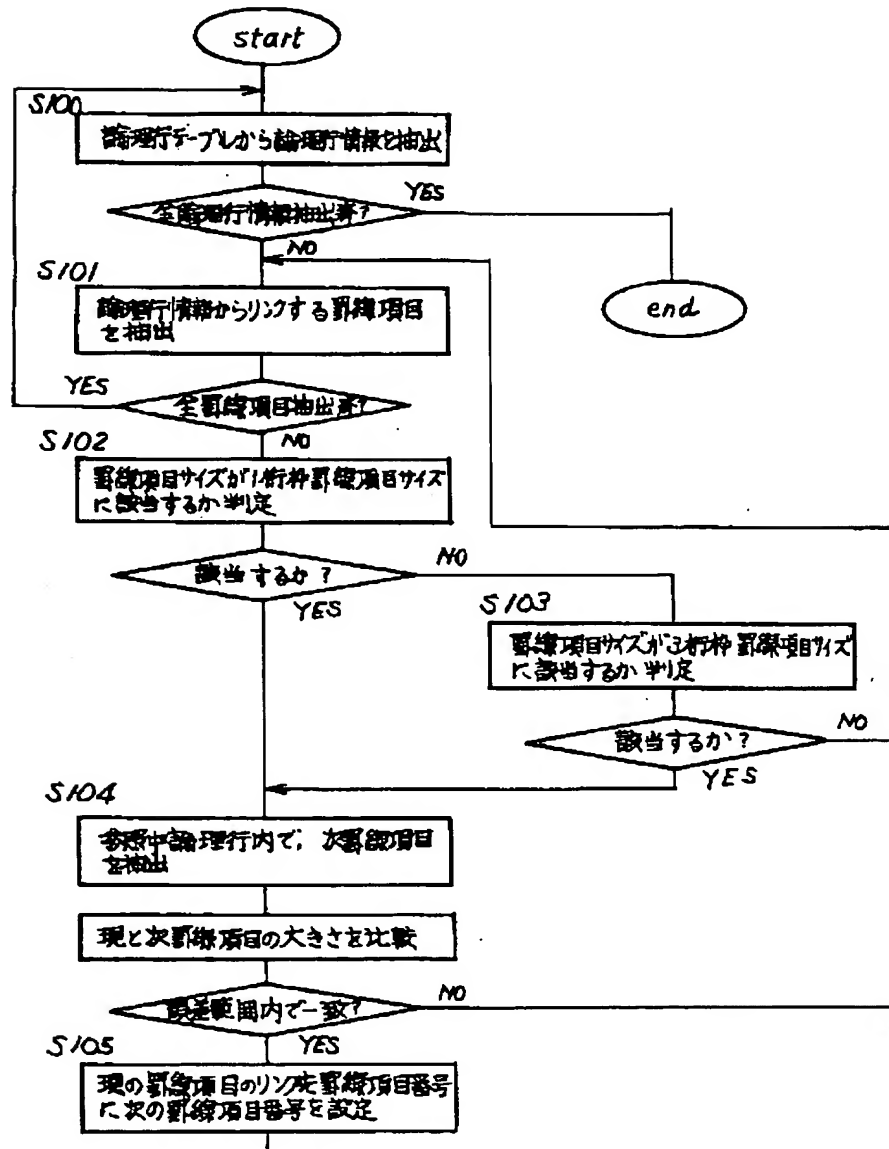
見出し名称: 金融人カネ
認識結果データ数: 1
認識結果コード長: 9
認識結果データアドレス
見出し名称: 金融人カネ
認識結果データ数: 1
認識結果コード長: 6
認識結果データアドレス
見出し名称: 金融人カネ
認識結果データ数: 1
認識結果コード長: 10
認識結果データアドレス
見出し名称: 金融人カネ
認識結果データ数: 1
認識結果コード長: 10
認識結果データアドレス
見出し名称: 金融人カネ
認識結果データ数: 10
認識結果コード長: 5
認識結果データアドレス

認識結果データ領域 / 202

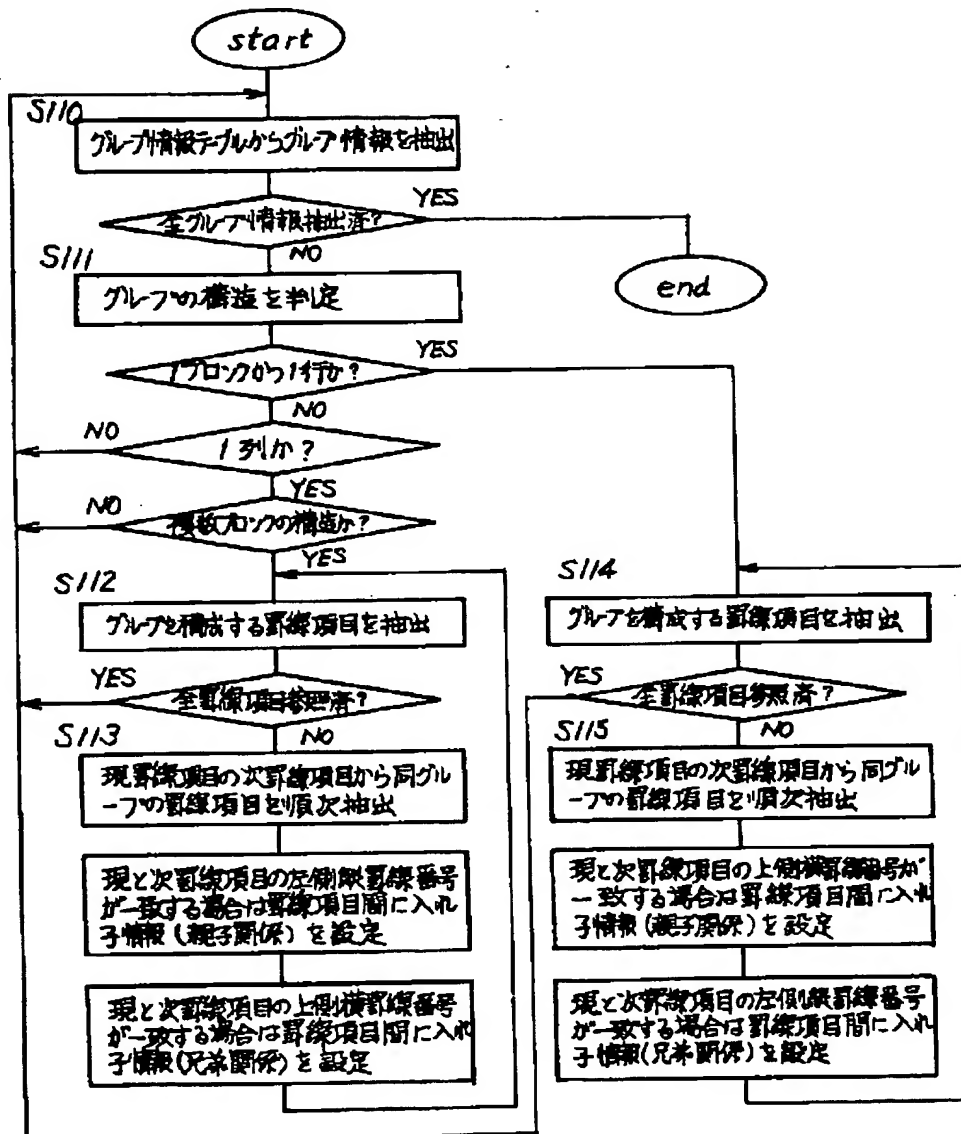
フジツタ
富士通 太郎
平成10年5月20日
平成10年5月20日
いろは銀行
みほに銀行
はにほ銀行
にほへ銀行
ほへち銀行
へちち銀行
とちり銀行
ちりぬ銀行
りぬる銀行
ぬるる銀行

【図 7 2】

1桁, 3桁罫線の入れ子解析処理フロー図



主表以外の入れ子解析処理フロー図



【図75】

入れ子解析後のテーブル説明図 (その1)

経路項目情報テーブル

RN	GN	BN	IN	RO	URN	LRN	ERN	RRN	IN					HAA
									INF	KN	CRN	RLN	AN	
1	1	1	1	1	2				ON	0	2	0	0	経路項目277172
2	1	1	1	1	4			3	ON	1	4	0	0	経路項目377172
3	1	1	1	1	5	2			ON	1	5	1	0	経路項目477172
4	1	1	1	1	2			5	ON	2	0	0	0	経路項目577172
5	1	1	1	1	3			4	ON	2	0	1	0	

論理行情報テーブル

LN	SC	EC	GN	BN	RIN	LL	ULN	LLN	ELN	RRN	HLA
1	X1,Y1	X2,Y2	1	1	5	1	1	3	1	4	経路項目1の77172

【図76】

入れ子解析後のテーブル説明図 (その2)

論理列情報テーブル

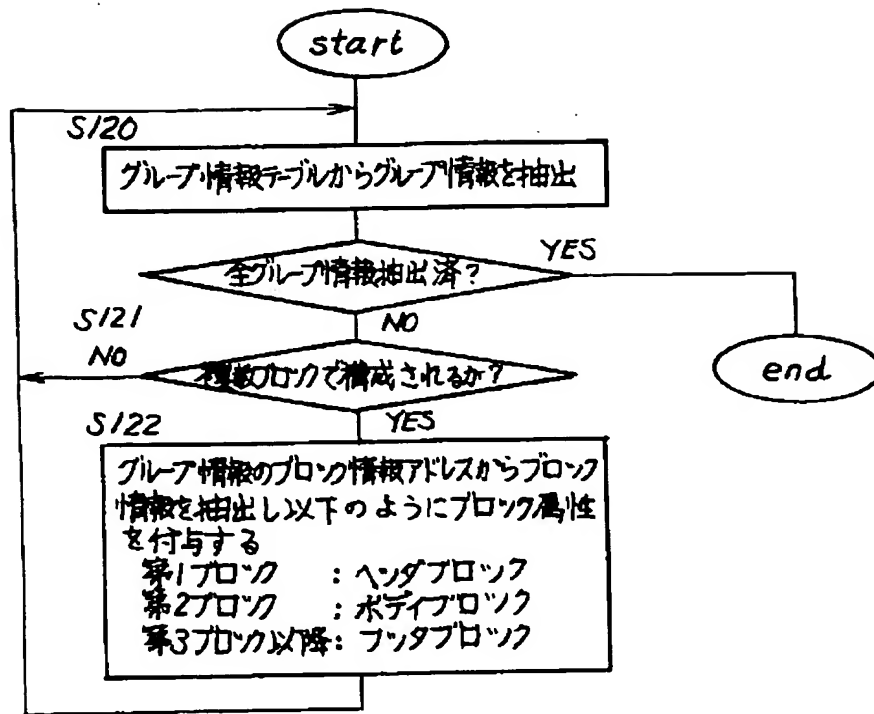
RO	SC	EC	GN	BN	RIN	ULN	LLN	ELN	RRN	HLA
1	X1,Y1	X2,Y2	1	1	5	1	3	1	4	経路項目1の77172
2	X3,Y3	X4,Y4	1	1	2	6	10	0	10	経路項目2の77172
3	X5,Y5	X6,Y6	1	1	2	6	10	10	11	経路項目3の77172

ブロック情報テーブル

BN	SC	EC	GN	RIN	LN	RO	HRA	HAA	ULN	LLN	LVN	RYN
1	X1,Y1	X2,Y2	1	5	1	1	経路項目1の77172	経路項目1の77172	1	3	1	4

【図 77】

属性付与処理フロー図



【図 78】

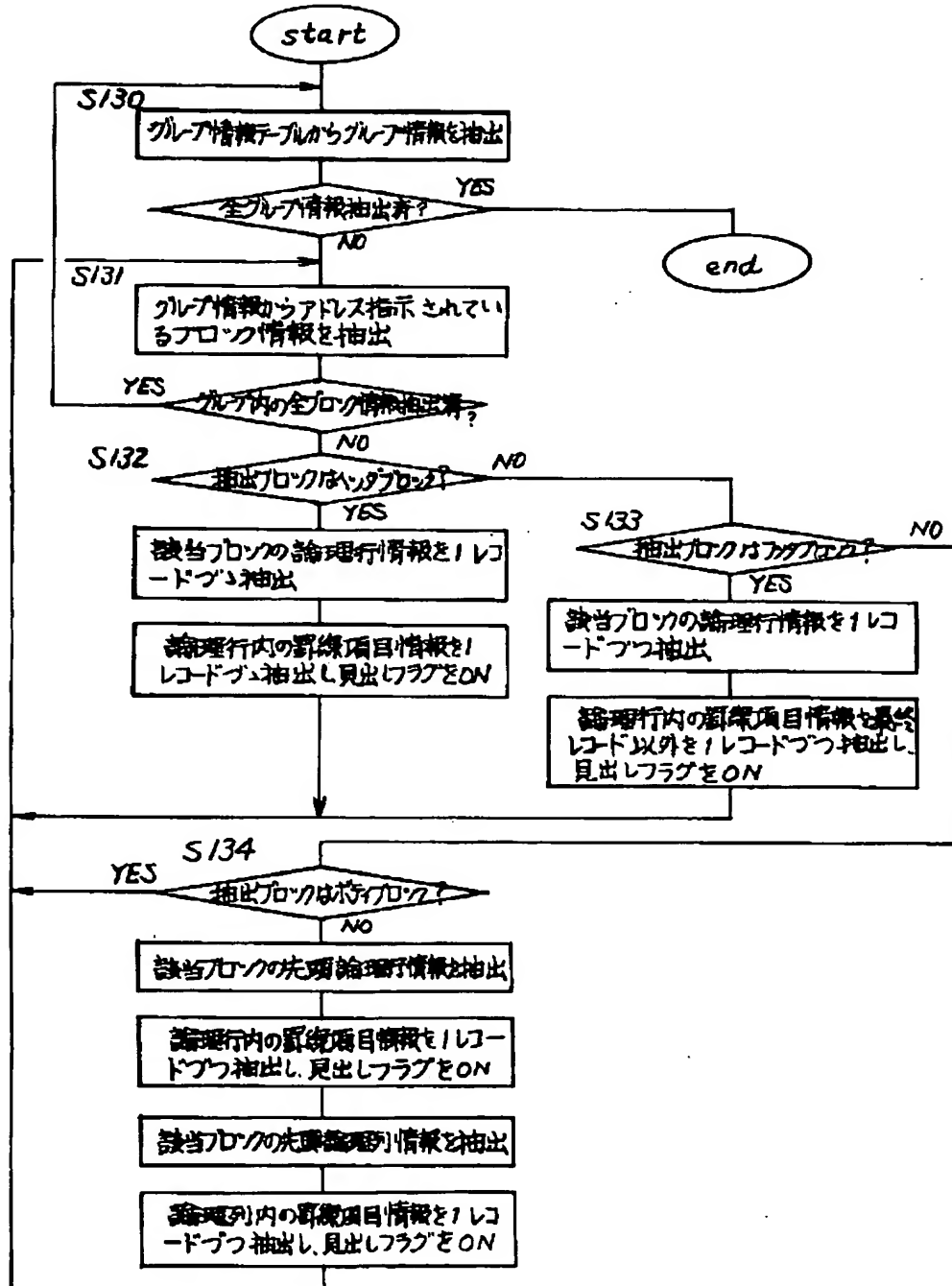
属性付与後のテーブル説明図

ブロック情報テーブル

BN	SC	EC	GN	RIN	LN	RO	HRA	HAA	BZ
1	X1.Y1	X2.Y2	1	5	1	3	0000000000000000	0000000000000000	0
1	X1.Y1	X2.Y2	2	4	2	2	0000000000000000	0000000000000000	0
1	X3.Y3	X4.Y4	3	7	1	5	0000000000000000	0000000000000000	1
2	X5.Y5	X6.Y6	3	140	10	5	0000000000000000		2
1	X7.Y7	X8.Y8	4	20	2	10	0000000000000000	0000000000000000	3

【図 79】

フラグ設定処理フロー図



【図 80】

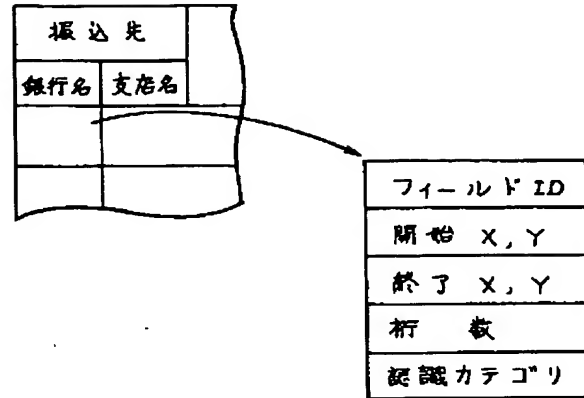
フラグ設定後のテーブル説明図

振込項目情報テーブル1102

RN	GN	BN	IN	RO	INF	KN	CRN	RIN	AN	MF	HAA
1	1	1	1	1	ON	0	2	0	0	1	振込項目2777777
2	1	1	1	1	ON	1	4	0	0	1	振込項目3777777
3	1	1	1	1	ON	1	5	1	0	1	振込項目4777777
4	1	1	1	1	ON	2	0	0	0	1	振込項目5777777
5	1	1	1	1	ON	2	0	1	0	1	
6	2	1	1	1						1	振込項目8777777
7	2	1	1	2						0	振込項目9777777
8	2	1	2	1						1	
9	2	1	2	2						0	
10	3	1	1	1	ON	0	11	0	0	1	振込項目1177777
17	3	2	1	1	ON	0	0	0	0	1	振込項目1877777
18	3	2	1	1	ON	0	0	1	0	1	振込項目1377777
157	4	1	1	1						1	振込項目1677777
158	4	1	1	2	ON	0	0	0	159	1	振込項目1577777

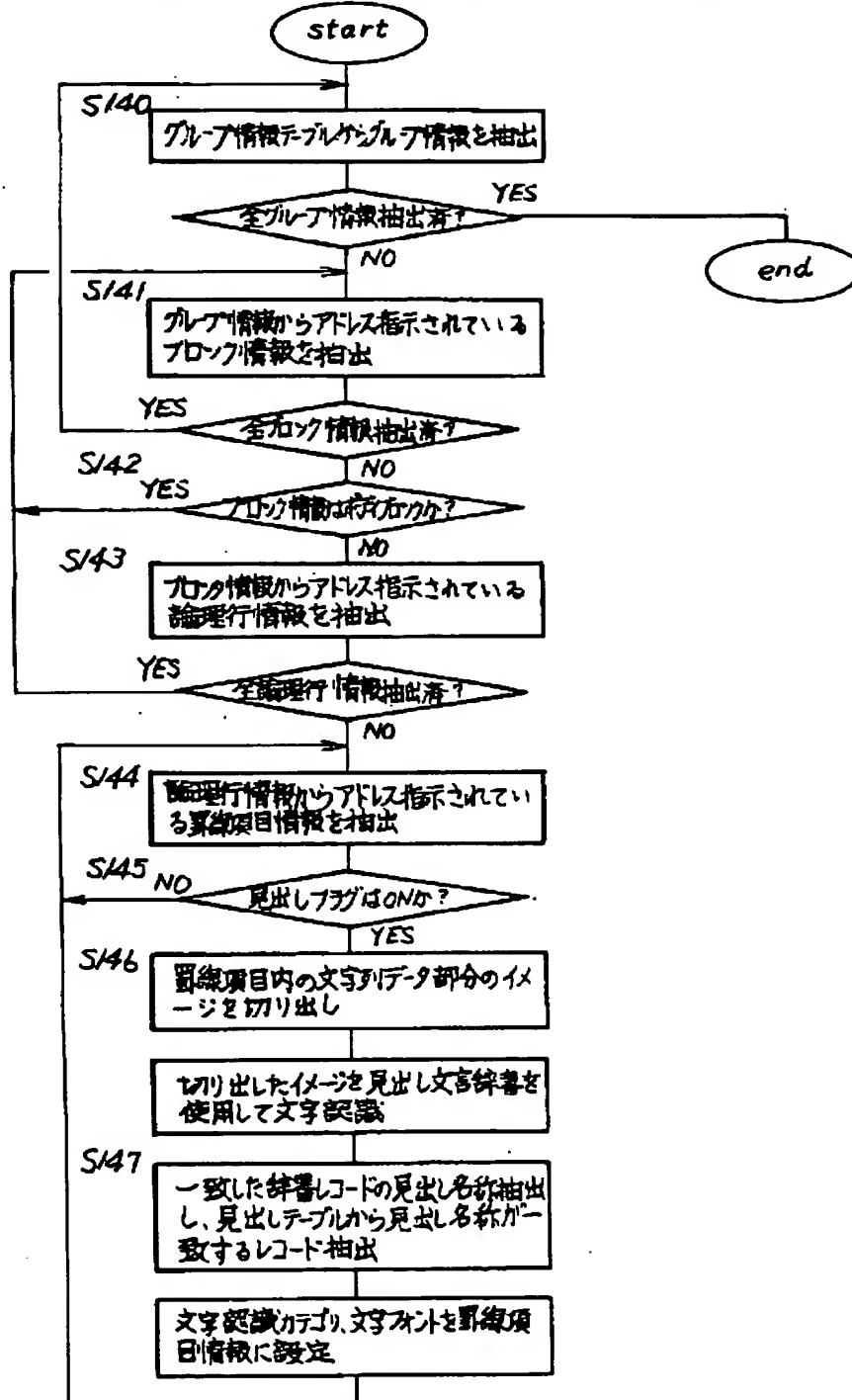
【図 87】

従来技術の説明図



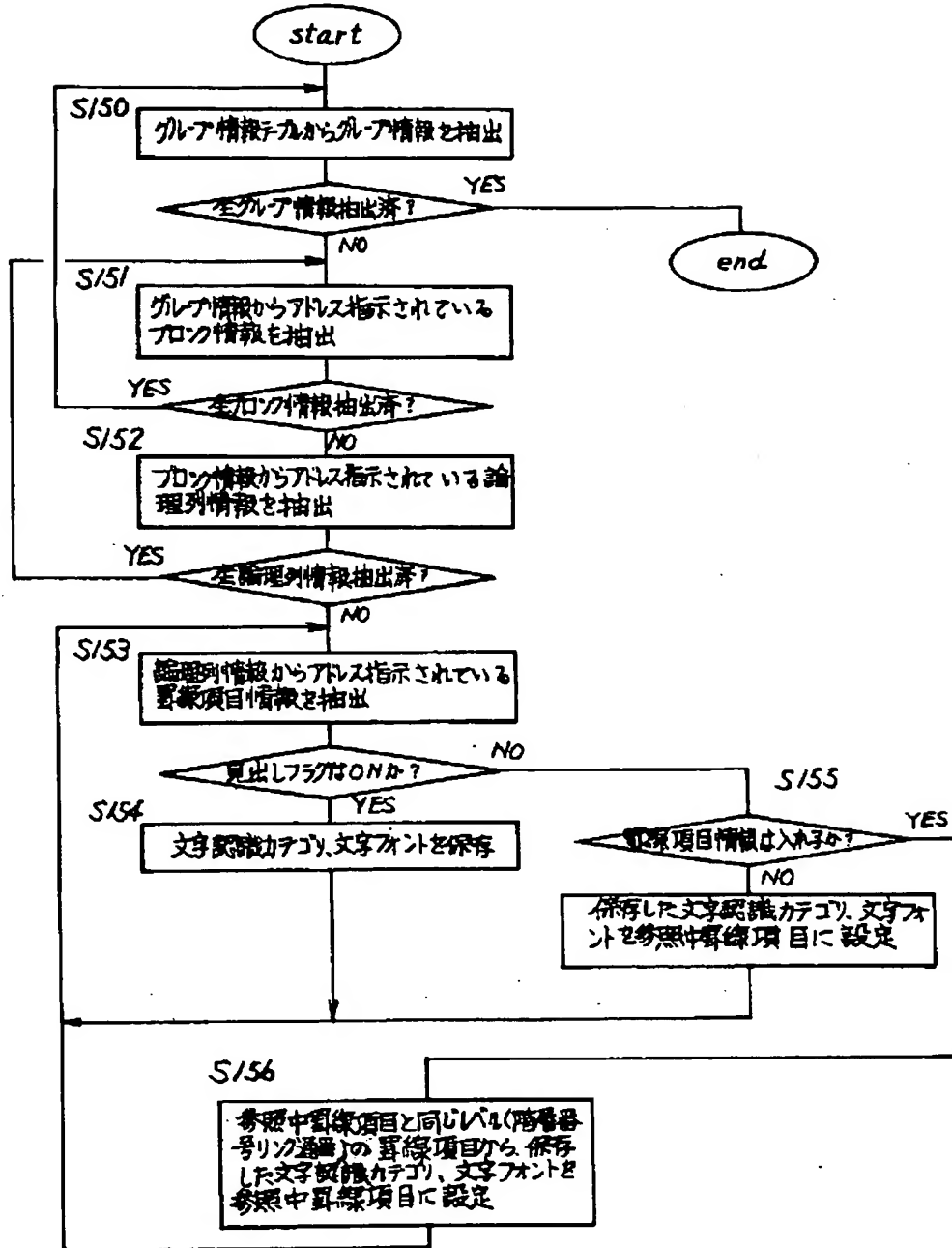
【図 81】

見出し認識処理フロー図



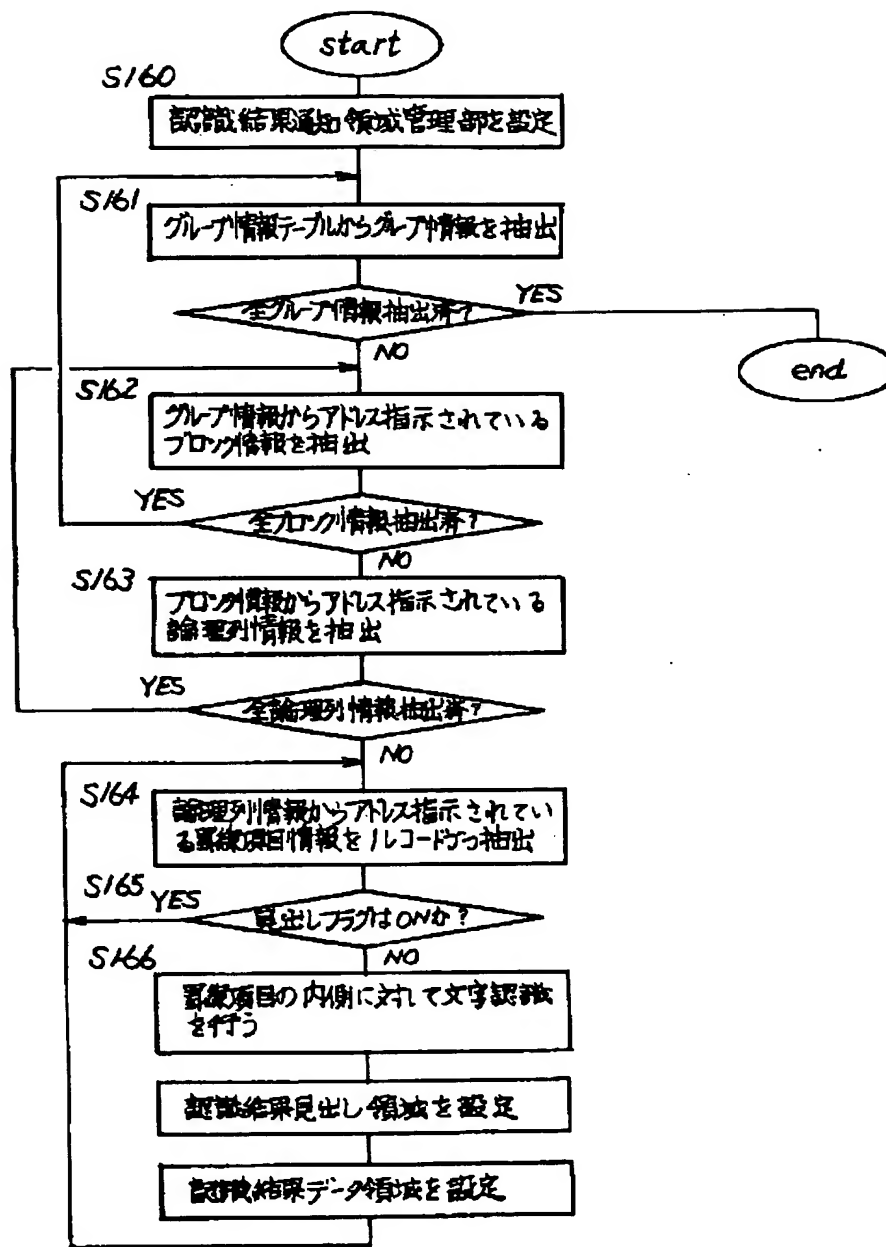
【図 83】

属性設定処理フロー図



【図 85】

文字認識処理フロー図



【手続補正書】

【提出日】平成11年8月6日(1999. 8. 6)

【補正方法】変更

【手続補正1】

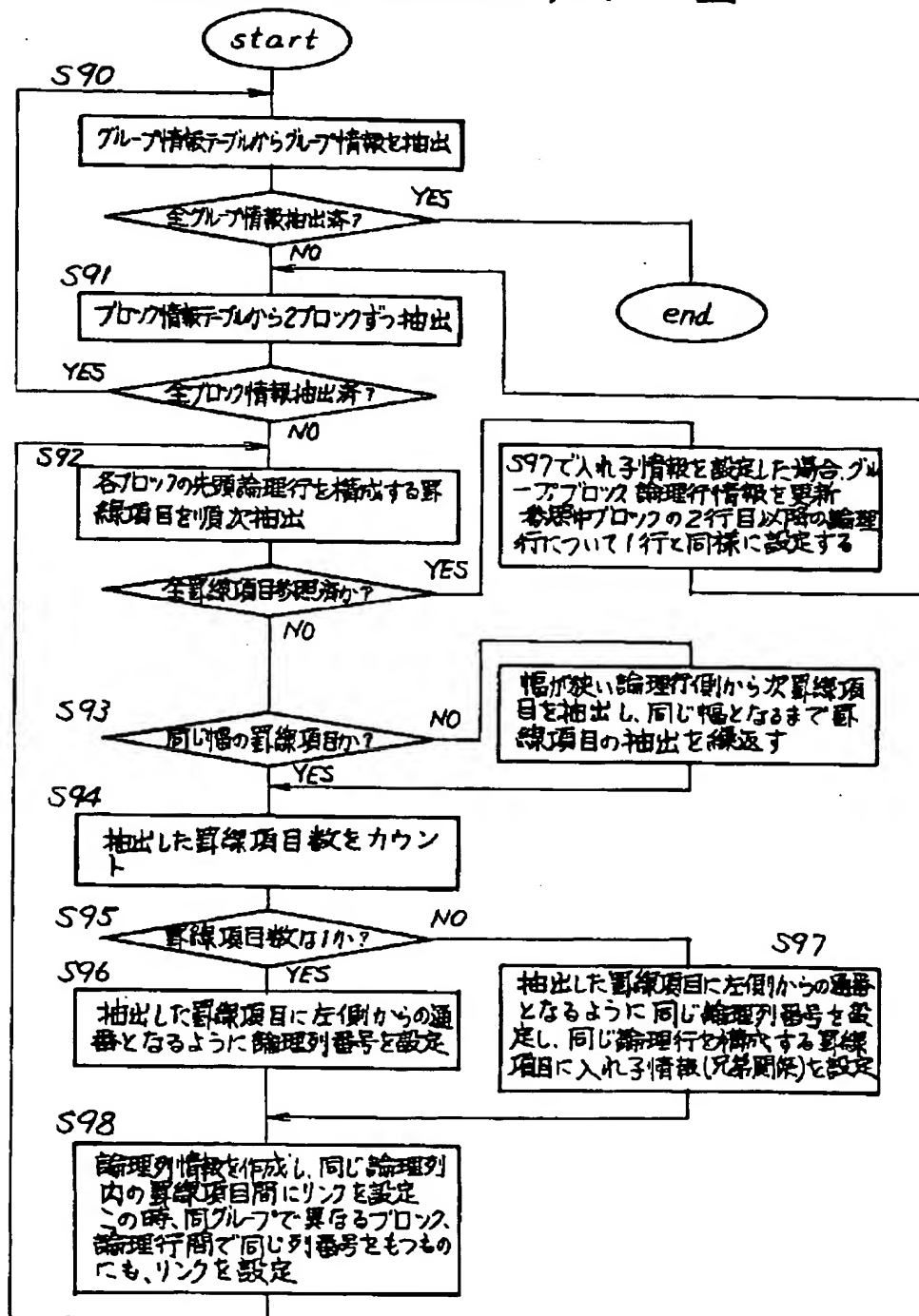
【補正内容】

【補正対象書類名】図面

【図68】

【補正対象項目名】図68

論理列判定処理フロー図



フロントページの続き

(72)発明者 千葉 亘一
 神奈川県川崎市中原区上小田中4丁目1番
 1号 富士通株式会社内

(72)発明者 金元 浩一
 神奈川県川崎市中原区上小田中4丁目1番
 1号 富士通株式会社内

(72)発明者 矢吹 真紀
神奈川県川崎市中原区上小田中 4 丁目 1 番
1 号 富士通株式会社内

(72)発明者 勝又 裕
神奈川県川崎市中原区上小田中 4 丁目 1 番
1 号 富士通株式会社内
F ターム(参考) 5B029 CC26 EE12
5L096 BA18 FA03 FA16 FA79 HA13